

Genome-led Discovery of Novel RiPP Natural Products

Alicia H Russell

**John Innes Centre
Department of Molecular Microbiology**

A thesis submitted to the University of East Anglia for the degree of
Doctor of Philosophy.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Ribosomally synthesised and post-translationally modified peptides (RiPPs) are a structurally diverse class of natural product that display a range of clinically relevant bioactivities such as antimicrobial and anticancer. The full extent of RiPP biochemical complexity has not yet been fully explored. This is partly due to challenges associated with genome mining for RiPP biosynthetic gene clusters (BGCs), which is often hampered by poor detection of the short precursor peptides that are ultimately modified into the final metabolite. Microorganisms are therefore predicted to produce many more RiPPs than are currently known.

In this thesis, a novel RiPP genome mining tool, RiPPER, is employed to identify novel RiPP precursor peptides near YcaO-domain proteins, enzymes that catalyse various RiPP post-translational modifications including heterocyclisation and thioamidation. From this analysis, I report the identification of a novel and diverse family of RiPP BGCs present in over 230 species of Actinobacteria and Firmicutes.

A representative BGC from *Streptomyces albus* J1074 was characterised through cloning and expression of the pathway followed by genetic, metabolomic and structural studies. I thus report the identification of a novel RiPP, streptomidine, which contains a structurally rare amidine ring. The identification of this metabolite along with bioinformatic analysis of homologous pathways suggests that amidine-containing RiPPs might be widespread in nature, where previously only two examples have been characterised. Bottromycin and klebsazolicin are both antibiotic RiPPs whose activity has been attributed to their amidine rings. Amidines might therefore represent an important structural feature for antibiotics that can be explored in the future.

Overall, these studies show that many more biochemically diverse natural products can be discovered through the use of targeted genome mining approaches, even in widely studied model organisms such as *Streptomyces albus* J1074. This represents an exciting prospect for the future of antimicrobial discovery.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Contents

Abstract	I
Figures and Tables	IX
Table of Figures	IX
Table of Tables	XII
Abbreviations	XIII
Acknowledgements	XVI
Research Outputs	XVII
Foreword	XVIII
Chapter 1: Introduction	1
1.1. Natural Products	1
1.1.1. Microbial Natural Products	1
1.1.2. A History of Microbial Natural Products as Antibiotics	2
1.1.3. The <i>Streptomyces</i> Genus	3
1.1.3.1. <i>Streptomyces</i> development	3
1.1.3.2. Natural products from <i>Streptomyces</i>	4
1.1.4. Natural Products from other Kingdoms of Life	5
1.1.4.1. Fungal natural products	5
1.1.4.2. Archaeal natural products	5
1.1.4.3. Animal natural products	5
1.1.4.4. Plant natural products	5
1.1.5. Antimicrobial Resistance	6
1.1.6. Approaches to Natural Products Discovery	7
1.1.6.1. Exploring new environments	8
1.1.6.2. Novel isolation methods	8
1.1.6.3. Direct analysis of environmental samples	9
1.1.7. Natural Product Classes	9
1.1.7.1. Polyketides	9
1.1.7.2. Terpenes	10
1.1.7.3. Nonribosomal peptides	10
1.1.7.4. Ribosomally synthesised and post-translationally modified peptides (RiPPs)	11
1.2. RiPPs	12
1.2.1. General Biosynthetic Logic	12
1.2.2. RiPP Subclasses	13
1.2.2.1. Linear azol(in)e-containing peptides (LAPs)	13
1.2.2.2. Thiopeptides	13
1.2.2.3. Cyanobactins	13

1.2.2.4. Bottromycins	14
1.2.2.5. Thioamitides	15
1.2.2.6. Lanthipeptides	15
1.2.2.7. Sactipeptides	16
1.2.2.8. Linaridins	16
1.2.2.9. Lanthidins	17
1.2.2.10. Lasso peptides	17
1.2.2.11. Streptide and darobactin	18
1.2.2.12. Proteusins	19
1.2.2.13. Other bacterial RiPPs	19
1.2.2.14. Fungal RiPPs	19
1.2.2.15. Plant RiPPs	20
1.2.3. Biosynthesis of Post-Translational Modifications in RiPPs	21
1.2.3.1. Understanding RiPP biosynthesis	21
1.2.3.2. Oxazol(in)e and thiazol(in)e rings	21
1.2.3.3. Thioamide bonds	22
1.2.3.4. Amidine rings	23
1.2.3.5. Methylations	23
1.2.3.6. Lysine-to-tryptophan cross-linking	24
1.2.3.7. Lanthionine bonds	24
1.2.3.8. Epimerisation	25
1.2.3.9. Removal of leader peptide	27
1.2.3.10. Biosynthesis of cyclic plant RiPPs	28
1.3. Thesis Aims	29
1.3.1. Importance of RiPP Discovery	29
1.3.2. Gaps in Current Understanding	29
1.3.3. Aims and Objectives of Thesis	29
1.3.4. Thesis Outline	29
Chapter 2: Genome Mining for Novel RiPP Gene Clusters	30
2.1. Introduction	30
2.1.1. Genome Mining and the Genomic Era	30
2.1.2. Genome Mining for RiPP Biosynthetic Gene Clusters	32
2.1.2.1. Analysis of whole genomes to identify RiPP biosynthetic gene clusters	32
2.1.2.2. Mass spectrometry-guided mining tools for RiPPs	34
2.1.2.3. Bespoke approaches to RiPP genome mining	37
2.1.2.4. Application of recently developed genome mining tools to discover novel RiPPs	38
2.1.3. RiPPER: a Novel Gene-led Genome Mining Tool for RiPPs	40

2.1.3.1. RiPPER workflow	40
2.1.3.2. RiPPER-led discovery of the thiovarsolins.....	41
2.2. Chapter Aims	43
2.3. Results and Discussion	44
2.3.1. Comparative Analysis of Genome Mining Outputs from the Genome of <i>Streptomyces albus</i> J1074.....	44
2.3.2. Using RiPPER to Identify Novel RiPP Gene Clusters Associated with YcaO-domain Proteins.....	46
2.3.2.1. Analysis of Actinobacterial YcaO-domain proteins.....	46
2.3.2.2. Retrieval and analysis of precursor peptides associated with YcaO-domain proteins.....	48
2.3.2.3. Phylogenetic analysis of associated YcaO-domain proteins	50
2.3.3. Bioinformatic Analysis of a Novel RiPP Precursor Peptide Family	51
2.3.3.1. Sequence alignment and motif identification.....	51
2.3.3.2. Distribution of identified precursor peptides in nature	52
2.3.3.3. Relationship between newly identified precursor peptides and their corresponding YcaO-domain proteins	55
2.3.4. Comparison of Gene Cluster Architectures	56
2.3.4.1. MultiGeneBlast analysis	56
2.3.4.2. Pathway regulation.....	59
2.3.4.3. Comparison of gene cluster architecture with the associated YcaO and precursor phylogeny.....	60
2.4. Chapter Summary	61
Chapter 3: Cloning & Expression of a Novel RiPP Gene Cluster from <i>Streptomyces albus</i> J1074.....	62
3.1. Introduction	62
3.1.1. Genetic Manipulation of Natural Product Biosynthetic Gene Clusters	62
3.1.2. Cloning Natural Product Biosynthetic Gene Clusters	63
3.1.3. <i>Streptomyces albus</i> J1074 as a Model Organism	63
3.2. Chapter Aims	64
3.3. Results and Discussion	65
3.3.1. Overview of the <i>Streptomyces albus</i> J1074 Biosynthetic Gene Cluster.....	65
3.3.2. TAR Cloning of the <i>S. albus</i> Biosynthetic Gene Cluster	67
3.3.2.1. TAR cloning design	67
3.3.2.2. Construction of pCAP03-derived capture vector (pSalbCAP)	67
3.3.2.3. Spheroplast transformation and screening of cluster capture	68
3.3.3. Heterologous Expression and Metabolomic Screening	71
3.3.3.1. Deletion of precursor peptide gene from pCAPSalbC.....	71
3.3.3.2. Heterologous expression.....	71
3.3.3.3. Untargeted metabolomic screening	72

3.3.4. Insertional Disruption of the <i>S. albus</i> J1074 Biosynthetic Gene Cluster	76
3.3.4.1. Construction of disruption vector	76
3.3.4.2. Conjugation of <i>S. albus</i> J1074 with pKCΔ <i>amiCD</i>	77
3.3.4.3. Metabolomic screening of <i>S. albus</i> J1074 and <i>S. albus</i> -pKCΔ <i>amiCD</i> ..	77
3.3.5. Comparing Metabolites Produced by the Native and Heterologous Hosts....	80
3.3.5.1. GNPS networking analysis	80
3.3.6. Gene Deletions from the TAR Cloned <i>S. albus</i> J1074 Gene Cluster	82
3.3.6.1. Deletion of biosynthetic genes from pCAPSalbC	82
3.3.6.2. Metabolomic screening of gene deletion mutants	82
3.3.7. Genetic Complementation of Gene Deletions	84
3.3.7.1. Construction of gene complementation vectors	84
3.3.7.2. Metabolomic screening of genetically complemented mutants	84
3.4. Chapter Summary	85
Chapter 4: Investigating the Structure, Biosynthesis & Activity of a Novel RiPP	87
4.1. Introduction	87
4.1.1. Structural Characterisation of Natural Products	87
4.1.2. Biological Characterisation of Natural Products	88
4.2. Chapter Aims	89
4.3. Results and Discussion	90
4.3.1. Core Peptide Prediction and Structural Hypotheses.....	90
4.3.1.1. Core peptide prediction	90
4.3.1.2. Initial structural hypothesis	90
4.3.1.3. Molecular formula prediction.....	91
4.3.2. Purification of Metabolite.....	92
4.3.2.1. Small-scale purification trials	92
4.3.2.2. Large-scale purification	93
4.3.3. Structural Elucidation	96
4.3.3.1. Nuclear magnetic resonance (NMR) analysis.....	96
4.3.3.2. Structural confirmation using a fluorescamine-binding assay.....	101
4.3.3.3. Determination of stereochemistry of streptamidine	102
4.3.3.4. Spontaneous isomerisation of streptamidine	103
4.3.3.5. MicroED analysis of streptamidine.....	104
4.3.3.6. Structures of streptamidine-related pathway metabolites.....	104
4.3.3.7. Is streptamidine the final pathway product?	107
4.3.4. Investigating the Biosynthesis of Streptamidine	108
4.3.4.1. Biosynthetic shunt metabolites	108
4.3.4.2. Biosynthetic mechanism of amidine formation.....	110
4.2.4.3. Dissecting the roles of RiPP tailoring enzymes in streptamidine biosynthesis	111

4.2.4.4. Mutational analysis of streptomidine core peptide	115
4.3.5. Prevalence of Streptomidines in Nature	117
4.3.6. Biological Role of Streptomidines in Nature	120
4.3.6.1. Antimicrobial assays.....	120
4.3.6.2. Metal-binding assays.....	122
4.3.6.3. Developmental assays	122
4.4. Chapter Summary	123
Chapter 5: General Discussion & Future Work.....	124
5.1. Genome Mining for RiPPs	124
5.1.1. Streptomidine Discovery	124
5.1.2. Future Work with RiPPER and YcaO-containing Pathways.....	124
5.2. Pathway Cloning and Engineering.....	125
5.2.1. Cloning and Genetic Manipulation	125
5.2.2. Synthetic Biology	125
5.3. Structure and Biosynthesis of Streptomidine	126
5.3.1. Importance of Amidines in Nature	126
5.3.2. Bioactivity of Streptomidine	127
5.3.3. Biosynthesis of Streptomidine	128
5.4. Conclusion	129
Chapter 6: Materials & Methods	130
6.1. Materials	130
6.1.1. Strains	130
6.1.2. Plasmids	133
6.1.3. Chemicals and Media	134
6.1.4. Oligonucleotides	137
6.2. General Methods.....	141
6.2.1. <i>E. coli</i>	141
6.2.1.1. Growth and maintenance	141
6.2.1.2. Transformation	141
6.2.1.3. Triparental mating	142
6.2.1.4. Plasmid isolation	142
6.2.2. <i>Streptomyces</i> Species	142
6.2.2.1. Growth and maintenance	142
6.2.2.2. Genomic DNA extraction	142
6.2.2.3. Conjugation	143
6.2.3. Cloning and Sequencing.....	144
6.2.3.1. Polymerase Chain Reaction (PCR)	144
6.2.3.2. Agarose gel electrophoresis	147
6.2.3.3. Purification of DNA from agarose gel.....	147

6.2.3.4. Purification of DNA from PCR mixtures	147
6.2.3.5. DNA digestions.....	147
6.2.3.6. Ligations.....	147
6.2.3.7. Sequencing	147
6.2.4. Mass Spectrometry	148
6.2.4.1. Preparation of mass spectrometry samples from culture extracts	148
6.2.4.2. Standard LC-MS analysis.....	148
6.2.4.3. High resolution LC-MS ² analysis	148
6.3. Bioinformatics and Genome Mining	149
6.3.1. Analysis of YcaO-domain Proteins.....	149
6.3.1.1. Retrieval of YcaO-domain proteins from Actinobacteria.....	149
6.3.1.2. Sequence alignments and phylogenetic tree building	149
6.3.2. Genome Mining	149
6.3.2.1. RiPPER	149
6.3.2.2. Whole genome comparative analysis	149
6.3.3. Analysis of Precursor Peptides	149
6.3.3.1. Precursor peptide analysis	149
6.3.3.2. Sequence alignments	150
6.3.3.3. Identification of motifs.....	150
6.3.3.4. Evolutionary networking	150
6.3.3.5. Comparison of gene cluster architectures.....	150
6.3.4. Analysis of <i>S. albus</i> J1074 Biosynthetic Gene Cluster	150
6.4. Characterising the <i>S. albus</i> J1074 Gene Cluster	151
6.4.1. TAR Cloning	151
6.4.1.1. Construction of capture vector.....	151
6.4.1.2. DNA digestion	151
6.4.1.3. Spheroplast transformation.....	152
6.4.1.4. Screening for captured gene cluster	153
6.4.2. Heterologous Expression	153
6.4.3. Gene Cluster Disruptions	153
6.4.4. Gene Deletions	154
6.4.5. Complementations	155
6.4.6. Core Peptide Mutations	155
6.4.7. Fermentation and Metabolomic Screening.....	156
6.4.7.1. Production cultures.....	156
6.4.7.2. Untargeted metabolomic analysis.....	156
6.4.7.3. Metabolite networking.....	156
6.5. Purification and Characterisation of Streptamidine	157
6.5.1. Purification of Metabolite.....	157

6.5.1.1. Large scale cultures	157
6.5.1.2. Liquid extraction	157
6.5.1.3. Solid phase extraction	157
6.5.1.4. High performance liquid chromatography (HPLC)	157
6.5.2. Structural Elucidation	158
6.5.2.1. NMR (Nuclear Magnetic Resonance)	158
6.5.2.2. Fluorescamine-binding assay	158
6.5.2.3. Marfey's analysis.....	158
6.5.3. Bioassays	159
6.5.3.1. Antimicrobial assays.....	159
6.5.3.2. Metal binding assays.....	159
6.5.4. MASST Analysis	159
References	160
Appendices.....	197

Figures and Tables

Table of Figures

Figure 1: Examples of natural products produced by diverse organisms	6
Figure 2: Examples of metabolites from different natural product classes	11
Figure 3: General biosynthetic logic of RiPPs.....	12
Figure 4: Chemical structures of RiPPs containing (methyl)oxazole and thiazole heterocycles	14
Figure 5: Chemical structure of thioalbamide	15
Figure 6: Chemical structures of nisin A and cypemycin	17
Figure 7: Chemical structure of microcin J25.....	18
Figure 8: Chemical structures of streptide and darobactin	19
Figure 9: Mechanisms of YcaO-domain protein-mediated catalysis.....	22
Figure 10: Mechanism of (methyl)lanthionine bond formation.....	25
Figure 11: Mechanisms of epimerisation	26
Figure 12: Cumulative number of prokaryotic genomes deposited in the NCBI database	31
Figure 13: Examples of RiPP natural products discovered by genome mining	39
Figure 14: RiPPER output of an annotated microbial genome viewed with Artemis	41
Figure 15: Chemical structures of thiovarsolins A-D.....	42
Figure 16: Genome mining outputs from analysis of the <i>S. albus</i> J1074 genome	45
Figure 17: EFI-EST analysis of Actinobacterial YcaO-domain proteins	47
Figure 18: Precursor peptide networks identified from RiPPER analysis	49
Figure 19: Phylogenetic tree of standalone YcaO-domain proteins from Actinobacteria	50
Figure 20: MEME analysis of peptide sequences from novel RiPP precursor network 1	51
Figure 21: Comparison of precursor network 1 with corresponding sequence motifs.....	52
Figure 22: EGN networking analysis of precursor peptides within network 1.....	54
Figure 23: Phylogenetic tree of YcaO-domain proteins associated with precursor network 1 ..	55
Figure 24: Example BGCs homologous to that of <i>S. albus</i> J1074.....	57
Figure 25: Example biosynthetic gene clusters that are related to that of <i>S. albus</i> J1074.....	59
Figure 26: Comparison of YcaO phylogeny, gene clusters and precursor sub-networks	60
Figure 27: Putative RiPP biosynthetic gene cluster (<i>ami</i>) from <i>S. albus</i> J1074	66
Figure 28: PCR screen of Image of plasmids isolated after Gibson assembly of pSalbCAP	68
Figure 29: PCR screen of pooled colonies following TAR cloning.....	69
Figure 30: PCR screen of individual colonies following pooled batch screening.....	69
Figure 31: Analytical restriction digest of pCAPSalbC construct	70
Figure 32: Sequence alignment showing deletion of precursor peptide from pCAPSalbC	71
Figure 33: Comparison of base peak chromatograms from LC-MS analysis of <i>S. coelicolor</i> M1146-pCAPSalbC and <i>S. coelicolor</i> M1146-pCAPSalbCΔ <i>amiA</i>	72
Figure 34: Profiling Solution analysis of metabolites produced by <i>S. coelicolor</i> M1146-pCAPSalbC and <i>S. coelicolor</i> M1146-pCAPSalbCΔ <i>amiA</i>	73
Figure 35: Comparison of extracted ion chromatograms of pathway-associated metabolites from <i>S. coelicolor</i> M1146-pCAPSalbC and <i>S. coelicolor</i> M1146-pCAPSalbCΔ <i>amiA</i>	74
Figure 36: High-resolution LC-MS ² data for pathway-related metabolites.....	75
Figure 37: PCR screen of pKCΔ <i>amiCD</i> disruption construct.....	76
Figure 38: PCR screen of <i>S. albus</i> J1074-pKCΔ <i>amiCD</i> exconjugants.....	77
Figure 39: Comparison of extracted ion chromatograms of pathway-associated metabolites from <i>S. albus</i> J1074 and <i>S. albus</i> -pKCΔ <i>amiCD</i>	79
Figure 40: Comparison of MS ² data for <i>m/z</i> 324.16 [M+2H] ²⁺ from <i>S. coelicolor</i> M1146-pCAPSalbC <i>S. albus</i> J1074	80
Figure 41: GNPS molecular network of pathway-associated metabolites	81
Figure 42: LC-MS analysis of key biosynthetic gene deletions from pCAPSalbC.....	83
Figure 43: Heat map of metabolite production in pathway mutants	83
Figure 44: Genetic region TAR cloned from <i>S. albus</i> J1074	84
Figure 45: LC-MS analysis of genetic complementations of deletions from pCAPSalbC	85
Figure 46: Structural hypotheses for the putative final pathway metabolite	91

Figure 47: Time course of pathway-associated metabolite production	92
Figure 48: LC-MS analysis of fractions obtained during purification of streptamidine.....	94
Figure 49: LC-MS analysis of combined HPLC fractions containing streptamidine	95
Figure 50: LC-MS analysis of pure streptamidine following final HPLC purification step	95
Figure 51: Diagnostic HMBC correlation for the amidine ring in streptamidine	97
Figure 52: Detailed correlation data for the structural elucidation of streptamidine.....	98
Figure 53: Structure of streptamidine in relation to the precursor peptide AmiA	98
Figure 54: Full proton spectrum for streptamidine recorded in DMSO-d ₆	100
Figure 55: Full carbon spectrum for streptamidine in DMSO-d ₆	100
Figure 56: LC-MS analysis of fluorescamine reaction with L-histidine and streptamidine	101
Figure 57: LC-MS analysis of Marfey's analysis of streptamidine	102
Figure 58: Stereochemistry of streptamidine	103
Figure 59: Comparison of ¹ H NMR spectra before and after isomerisation of streptamidine ..	104
Figure 60: Predicted structures of streptamidine-related metabolites	105
Figure 61: High-resolution LC-MS ² data for putative pathway-related metabolites	106
Figure 62: Predicted structures of streptamidine-related shunt metabolites	109
Figure 63: LC-MS analysis of <i>S. coelicolor</i> M1146-SalbCΔ <i>amiE</i> culture extract spiked with synthetic standard of N-acetylated LSA peptide	109
Figure 64: Structures of RiPPs containing amidine rings, thiazole heterocycles and oxazole heterocycles	110
Figure 65: Proposed mechanism of amidine ring formation in streptamidine.....	111
Figure 66: MEME analysis of all streptamidine-like YcaO-domain proteins	112
Figure 67: Sequence alignment of YcaO-domain proteins involved in biosynthesis of diverse RiPP post-translational modifications	113
Figure 68: Secondary structure alignment of AmiE with McbC	115
Figure 69: PCR screen of <i>E. coli</i> HME68-pCAPSsalbC colonies transformed with a mutant oligonucleotide	116
Figure 70: PCR screening of pCAPSsalbC core peptide mutants.....	116
Figure 71: Networks of streptamidine-like precursor peptides in relation to genetic architecture of corresponding biosynthetic gene clusters	118
Figure 72: Putative structures of streptamidine-like metabolites based on variations of core peptide sequences	118
Figure 73: MASST analysis of streptamidine.....	119
Figure 74: Bioassay plates generated during antimicrobial testing of streptamidine	120
Figure 75: Bioassay plate generated during co-culture experiments	121
Figure 76: Example of spot assay on high-salt agar.....	123
Figure 77: Chemical structures of amidine-containing metabolites	127
Figure 79: Genome mining output from RiPPMiner analysis of <i>S. albus</i> J1074 genome.....	197
Figure 79: Genome mining output from antiSMASH analysis of <i>S. albus</i> J1074 genome	197
Figure 80: Genome mining output from DeepRiPP analysis of <i>S. albus</i> J1074 genome	198
Figure 81: Genome mining output from BAGEL analysis of <i>S. albus</i> J1074 genome	198
Figure 82: Genome mining output from PRISM analysis of <i>S. albus</i> J1074 genome.....	199
Figure 83: Alignment of streptamidine-like precursor peptides.....	200
Figure 84: MEME output from analysis of streptamidine-like precursor peptides	202
Figure 85: MultiGeneBlast analysis of <i>S. albus</i> (<i>albidoflavus</i>) J1074 gene cluster	206
Figure 86: Vector map of pCAPSsalbC and 'Virtual gel' image of analytical digest.....	209
Figure 87: DNA alignment obtained from sequencing of pSalbCAP construct.....	209
Figure 88: Output from RiPP peptide mass calculator.....	212
Figure 89: Screenshot of output of Shimadzu formula prediction tool.	212
Figure 90: Fractions from Biotage purification that contain the target metabolite	213
Figure 91: UV chromatogram obtained from semi-prep HPLC purification of fractions containing target metabolite	213
Figure 92: UV chromatogram obtained following semi-prep HPLC purification of semi-pure material from previous HPLC purification step.....	214
Figure 93: 2D NMR correlations used to elucidate the structure of streptamidine	215
Figure 94: Structure of streptamidine with HMBC and COSY NMR correlations	216
Figure 95: 3D model of streptamidine	216
Figure 96: Chemical shifts for CH ₃ groups in streptamidine recorded in DMSO-d ₆	216

Figure 97: Amide and histidine proton signals for streptomidine recorded in DMSO-d ₆	217
Figure 98: Chemical shifts for CH and CH ₂ groups in streptomidine recorded in DMSO-d ₆ ...	217
Figure 99: Full HSQC spectrum for streptomidine in DMSO-d ₆	218
Figure 100: Full COSY spectrum for streptomidine in DMSO-d ₆	219
Figure 101: Full TOCSY spectrum for streptomidine recorded in DMSO-d ₆	220
Figure 102: Full HSQC-TOCSY spectrum for DMSO-d ₆	221
Figure 103: MS ² fragmentation data for predicted dehydrated LSA (<i>m/z</i> 272.1595) and predicted acetylated and dehydrated LSA (<i>m/z</i> 314.1711).....	222
Figure 104: Plasmid map of pKC1132	224
Figure 105: Plasmid map for pIJ10257	224
Figure 106: Plasmid map for pCAP03	225
Figure 107: Plasmid map for pIJ773Δ <i>oriT</i>	225

Table of Tables

Table 1: Details of predicted protein functions of the genes present in the <i>Streptomyces albus</i> J1074 biosynthetic gene cluster.....	66
Table 2: NMR chemical shift assignments for streptamidine in DMSO-d ₆	99
Table 3: List of microorganisms tested during bioactivity assays	121
Table 4: Strains used during study	130
Table 5: <i>Streptomyces</i> strains constructed during study	131
Table 6: Plasmids used during study	133
Table 7: Vectors constructed during study	133
Table 8: Antibiotics used during study	134
Table 9: Solutions used for TAR cloning	134
Table 10: Buffers used during study	134
Table 11: Media used during study	135
Table 12: Primers used during study	137
Table 13: Reaction conditions for <i>Streptomyces</i> colony PCR.....	144
Table 14: PCR cycling conditions for <i>Streptomyces</i> colony PCR	144
Table 15: Reaction conditions for <i>E. coli</i> colony PCR with G2 Taq polymerase	145
Table 16: PCR cycling conditions for <i>E. coli</i> colony PCR	145
Table 17: Reaction conditions for PCR cloning of genes from <i>Streptomyces</i> genomic DNA ..	146
Table 18: PCR cycling conditions for cloning of genes from <i>Streptomyces</i> genomic DNA	146
Table 19: Gibson assembly reaction mixtures	151
Table 20: DNA digestion for TAR cloning.....	152
Table 21: Ligation trial for pKC1132-based gene disruption constructs	154
Table 22: Metabolomic data of complete TAR clone (SalbC) expressed in <i>S. coelicolor</i> M1146 versus precursor peptide mutant (SalbCΔPP) and medium only (SM12)	210
Table 23: Metabolomic data of complete TAR clone (SalbC) expressed in <i>S. coelicolor</i> M1146 versus pathway mutants	210
Table 24: Metabolomic data of complete TAR clone (SalbC) expressed in <i>S. coelicolor</i> M1146 versus pathway mutants	211
Table 25: Metabolomic data of <i>S. albus</i> J1074 wild type versus pathway-disrupted mutant (<i>S. albus</i> YH mutant) and medium only (SM12)	211
Table 26: Accurate masses for streptamidine and related metabolites, obtained using a Waters Synapt G2Si	222
Table 27: Accurate masses for streptamidine and related metabolites, obtained using a Shimadzu IT-TOF.....	223

Abbreviations

ABC	ATP-binding cassette
Acc	1-aminocyclopropane-carboxylic acid
Aib	1-amino-isobutyric acid
Ala	alanine
antiSMASH	antibiotics and secondary metabolite analysis shell
ATP	adenosine triphosphate
AviCys	S-[(Z)-2-aminovinyl]-D-cysteine
BAGEL	bacteriocin genome mining tool
BARLEY	basic alignment of ribosomally encoded products locally
BGC	biosynthetic gene cluster
BLAST	basic local alignment search tool
BPC	base peak chromatogram
CDART	conserved domain architecture retrieval tool
CIPRES	cyberinfrastructure for phylogenetic research
CoA	coenzyme A
COSY	correlation spectroscopy
CRISPR	clustered regularly interspaced short palindromic repeats
CryoEM	cryogenic electron microscopy
CTX-M	cefotaxime-M
Da	dalton
Dha	dehydroalanine
Dhb	dehydrobutyrine
DMSO	dimethyl sulfoxide
DMAPP	dimethylallyl diphosphate
DNA	deoxyribonucleic acid
DNN	deep neural network
DUF	domain of unknown function
EGN	evolutionary gene and genome network
EST	enzyme similarity tool
ESBL	extended spectrum beta-lactamase
FMN	flavin mononucleotide
FPP	farnesyl diphosphate
GGPP	geranylgeranyl diphosphate
GNPS	global natural product social molecular networking
His	histidine

HMBC	heteronuclear multiple bond correlation
HMM	hidden markov model
HPLC	high performance liquid chromatography
HPSF	high purity salt free
HSEE	hypothetical structure enumeration and evaluation
HSQCed	heteronuclear single quantum correlation edited
HSQC-TOCSY	heteronuclear single quantum correlation-total correlation spectroscopy
iChip	isolation chip
iNOS	inducible nitric oxide synthase
IPP	isopentenyl diphosphate
iTOL	interactive tree of life
IT-TOF	ion trap-time of flight
Iva	isovaline
LAP	linear azol(in)e-containing peptide
LB	lysogeny broth
LC-MS	liquid chromatography-mass spectrometry
Leu	leucine
LiAc	lithium acetate
MALDI-TOF	matrix-assisted laser desorption/ionization-time of flight
MarR	multiple antibiotic resistance regulator
MASST	mass spectrometry search tool
(Me)Lan	(methyl)lanthionine
MEME	multiple em for motif elucidation
MicroED	microcrystal election diffraction
MRSA	methicillin-resistant <i>Staphylococcus aureus</i>
MS	mass spectrometry
MUSCLE	multiple sequence comparison by log-expectation
NAP	network annotation propagation
NCBI	national center for biotechnology information
NMR	nuclear magnetic resonance
NOESY	nuclear overhauser effect spectroscopy
NRPS	nonribosomal peptide synthetase
NPP	natural product peptidogenomics
ORF	open reading frame
PCR	polymerase chain reaction
PEG	polyethylene glycol
Pip	pipecolic acid

pHMM	profile hidden markov model
Phyre	protein homology/analogy recognition engine
pK_a	acid dissociation constant
PKS	polyketide synthase
PRISM	prediction informatics for secondary metabolomes
RAxML	randomised accelerated maximum likelihood
RiPP	ribosomally synthesised and post-translationally modified peptide
RiPPER	RiPP precursor peptide enhanced recognition
RNA	ribonucleic acid
RODEO	rapid ORF detection and evaluation online
rSAM	radical S-adenosyl-L-methionine
SAM	S-adenosyl-L-methionine
Ser	serine
SFM	soya flour mannitol
SM12	screening medium 12
SSN	sequence similarity network
TAR	transformation associated recombination
Thr	threonine
TOCSY	total correlation spectroscopy
TOMM	thiazole/oxazole-modified microcin
TMA	trimethylamine
tRNA	transfer ribonucleic acid
VRE	vancomycin-resistant Enterococci

Acknowledgements

Carrying out my PhD at the John Innes Centre has been an incredible experience thanks to all the wonderful people who have supported me along the way. I firstly want to thank Andy Truman for the opportunity to work on such a fun project, and for all the help and support that he has provided throughout. I am also grateful for the encouragement to learn and develop as a scientist, which has allowed me to gain a wide skill set across biology and chemistry. I would also like to thank my secondary supervisor Barrie Wilkinson for invaluable input into my research.

The Truman lab has been a joy to work in and members of the group (past and present) have taught me almost everything I have learnt over the past four years. I would particularly like to thank Natalia Miguel-Vior who has gone above and beyond to answer questions, teach me things in the lab and sit down with me for a chat whenever I was starting to feel a bit lost. Tom Eyles was a great source of support when I first arrived and Javier Santos-Aberturas has also been incredibly helpful throughout my PhD. Edward Hems and Rodney Lacret have been great chemistry teachers in the lab, and I would also like to thank Dave Widdick for being a source of endless wisdom and amusing anecdotes. Last but not least, Louis, Jonny and Alaster have been great lab mates who have always been willing to help me out. Beyond the Truman lab, the wider Mol Micro department have been amazing colleagues who have all helped create an unrivalled supportive family atmosphere.

I have shared my PhD journey with some lovely people who have helped make the last few years in Norwich a lot of fun: Charli, Aggie, Erin, Becky, Rob and Tom. The final few months of this PhD have been particularly strange due to the pandemic, and I would like to thank the Zoom quiz gang for helping to get me through this with a lot of laughs: Bex, Ali, Pip, Tom and Matt; as well as my corona support bubble, Tom. I would also like to thank Rich and Stuart for being at the end of the phone whenever I need to complain about how stressed I am. Finally I would like to thank my family for their support throughout my education and life in general: Mum, Dad, Dave and Laura.

Research Outputs

The research presented throughout this thesis will lead to published outputs.

Lead author

Streptamidine discovery

The story of streptamidine discovery (described in Chapters 2, 3, 4 and 5) is currently available as a pre-print on *bioRxiv*:

Russell AH, Vior NM, Hems ES, Lacret R, Truman AW. Discovery and characterisation of an amidine-containing ribosomally-synthesised peptide that is widely distributed in nature. *bioRxiv*. 2020; 2020.05.04.076059. <https://doi.org/10.1101/2020.05.04.076059>

RiPP genome mining (review)

A review of RiPP genome mining methods (summarised in Chapter 2) has been published in *Computational and Structural Biotechnology Journal*:

Russell AH and Truman AW. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Comput. Struct. Biotechnol. J.* 2020, 18, 1838–1851, <https://doi.org/10.1016/j.csbj.2020.06.032>

Co-author

Advances in actinomycete research (review)

General introductory sections (Chapter 1) include some of the themes summarised in a review of actinomycete research papers that has been published in *Microbiology*:

Prudence SMM, Addington E, Castaño-Espriu L, Mark DR, Pintor-Escobar L, Russell AH, McLean TC. *Microbiology* 2020, <https://doi.org/10.1099/mic.0.000944>

Characterisation of a novel siderophore

A research project led by Javier Santos-Aberturas is currently being carried out, to which I contributed by purifying and elucidating the structure of a novel siderophore from *Saccharopolyspora* sp. KY21.

Other outputs

ActinoBase

Some of the tools, resources and methods mentioned throughout this thesis have been uploaded onto <https://actinobase.org>, a website that I have contributed to as an editor.

Foreword

This thesis was written during the 2020 Covid-19 pandemic. In developed countries we have become familiar with a way of life where infectious disease is not at the forefront of our minds, but the present situation is a stark reminder of the destruction that infectious disease can cause on lives, health, jobs and economies, when no medical treatments are available.

In recent years, advances in medicine and public health measures have helped treat and reduce the prevalence of numerous infectious diseases. Widespread vaccination eradicated smallpox in 1979, and the second of three types of poliomyelitis virus was eradicated in 2019. Notably, one of the biggest breakthroughs in modern medicine was the discovery of antibiotics in the 1900s, which have played a huge role in treating deadly bacterial infections and enabling safe routine surgery to be carried out.

The majority of antibiotics used in the clinic are so-called 'natural products' produced by wide-ranging microbes which have been exploited for decades due to their biosynthetic talents. Unfortunately, due to recent overuse and misuse of antibiotics, bacteria are increasingly evolving resistance mechanisms that render these metabolites inactive. As a result, multidrug resistant 'superbugs' are starting to emerge, and antimicrobial resistance is a growing threat to public health that is predicted to cause millions more deaths in the future.

There is currently an urgent need for new antimicrobial metabolites to be discovered. Fortunately, the search from nature has by no means been exhausted. Recent analyses have shown that we are still scratching the surface of the huge biosynthetic capacity of microbes, indicating huge potential to discover new natural product antibiotics.

The importance of this thesis is to thus demonstrate how a combination of new and old scientific approaches can be applied to the mining, discovery and characterisation of novel natural products from bacteria, paving the way for future important discoveries in this field.

Chapter 1: Introduction

1.1. Natural Products

“The fact that certain species of bacteria have an inhibiting effect on the development of other species or varieties is too well known to need discussion.”

L. A. Rogers, 1928

1.1.1. Microbial Natural Products

For hundreds of years, microorganisms have been shown to exert inhibitory effects on other species (1,2). The causative metabolites produced by these organisms are known as ‘natural products’. As many of these metabolites exhibit antimicrobial or cytotoxic effects, a large number of microbial natural products have been utilised by the pharmaceutical industry to produce a plethora of human medicines (3). The development of antimicrobials, in particular, has had a significant effect in reducing the number of deaths attributed to infectious disease over the last century. More broadly, microbial natural products have also been used in agriculture as antibiotics, growth hormones and pesticides (4,5). Aside from antimicrobials, natural products can display other diverse biological functions, with roles in metal uptake, signalling, quorum sensing and development (6,7). Therefore, as well as important medicinal uses, natural products also represent an interesting aspect of microbial ecology. Due to their wide-ranging functions, and to distinguish them from primary metabolism, natural products can also be referred to as ‘secondary metabolites’ or ‘specialised metabolites’ (7).

The overarching theory of why microbes produce so many specialised metabolites with antimicrobial activity is that they act as a kind of defence mechanism, providing an advantage to microbes during warfare when fighting for nutrients and resources in their competitive natural environments. The protective role of natural products has been demonstrated in various ecological niches, such as in fungus-farming ants (8), beewolf diggerwasps (9) and southern pine beetles (10). In these cases, the insects are associated with antibiotic-producing actinomycete bacteria that protect mutualistic fungus against microbial predators. There are also theories that the antimicrobial activity of natural products is a secondary function to more important biological roles. For example, thiopeptide antibiotics were recently shown to stimulate biofilm formation in *Bacillus subtilis* (11), which is an important growth state for bacterial colonisation (12).

1.1.2. A History of Microbial Natural Products as Antibiotics

The start of the antibiotic era is often hallmarked by the discovery of penicillin from the fungus *Penicillium chrysogenum*, reported in 1929 (13) and later characterised in 1940 (14). However, the use of microbial extracts to treat disease has been alluded to much earlier on in history. For example, mouldy bread was listed as a remedy on the Eber's papyrus, the oldest preserved medical document from 1550 BC (15). In 1899, Emmerich and Löw reported that *Bacillus pycaneus* (now called *Pseudomonas aeruginosa*) could inhibit the growth of other microbes such as the causative agents of typhus and cholera (1). The extract produced by this bacterium was one of the first used as an antibacterial drug. Initially thought to be an enzyme, the 'pyocyanase' metabolite produced by *P. aeruginosa* were later confirmed to be pyocyanin and the 2-alkyl-4-quinolones, which are quorum sensing metabolites (16). Antibiosis between microbes was also reported by Pasteur in the early 1900s, who was one of the first scientists to propose that microbes might secrete metabolites to kill off other bacteria (17). A year before Fleming reported penicillin, a paper by Rogers from 1928 references the discovery of nisin, with the observation that *Lactococcus lactis* displays inhibitory activity against *Lactobacillus bulgaricus* (2). Nisin has now been used as a food preservative for over 50 years (18).

After the discovery of penicillin, several more antibiotics were subsequently isolated from microbes. In 1939, Dubos discovered tyrothricin, an antibiotic mixture of gramicidin and tyrocidine from the soil bacterium *Bacillus brevis* (19,20). Inspired by this, Selman Waksman started searching for further antibacterial metabolites from soil microbes by screening extracts for inhibitory activity. This approach, often referred to as 'the Waksman platform', led to the discovery of numerous antibiotics that are still in clinical use today, and earned Waksman the title of the 'father of antibiotics' (21). One of the first antibiotics discovered by Waksman, Bugie and Schatz was streptomycin, isolated from the soil bacterium *Streptomyces griseus* in 1943 (22). Streptomycin was another landmark antibiotic discovery as, unlike penicillin, it was active against *Mycobacterium tuberculosis*, whose infection was rife during and following the second world war. Following streptomycin, a number of other *Streptomyces*-derived antibiotics were isolated by Waksman's team throughout the 1940s-'60s, a period that has been coined the 'golden age' of antibiotic discovery (23,24).

Since Waksman's early work, it has been increasingly demonstrated that *Streptomyces* species and other members of the Actinobacteria phylum are particularly prolific producers of secondary metabolites. Natural products from *Streptomyces* species and related Actinobacteria are the focus of this thesis.

1.1.3. The *Streptomyces* Genus

The *Streptomyces* genus lies within the Streptomycetaceae family, Streptomycetales order and Actinobacteria class. Members of the Streptomycetales order are often referred to as actinomycetes, which represent a large group of filamentous and aerobic Gram-positive bacteria (25). The first actinomycete was described in 1875 by Cohn, who isolated a bacterium then named *Streptothrix foersteri* from a human tear duct (26). Later in 1943, Waksman and Henrici proposed a new genus classification for the group of actinomycetes that form characteristic branching hyphae, which included *S. foersteri*. This genus was named *Streptomyces*, which means “twisted fungus” (27). Today, the *Streptomyces* genus is particularly well studied, predominantly because members are talented producers of a range of secondary metabolites. Owing to this, streptomycetes have a relatively large genome for bacteria, typically eight megabases or larger (25), which forms a characteristic linear chromosome (28,29). Most *Streptomyces* bacteria live as saprophytes in the soil (30), from which many species have been selectively isolated. Actinobacteria make up 2-6% of the soil microbiome, as determined by 16S rRNA analysis of metagenomic samples (31). Many streptomycetes also inhabit other environmental niches such as fresh water, the desert, and marine environments including coral and sponges (32). The majority of *Streptomyces* species are non-pathogenic, but some act as animal or plant pathogens such as *Streptomyces scabies*, which is responsible for common potato scab disease (33). Streptomycetes can be classified based on phenotypic traits such as their spore-derived structures and the pigmentation of spores and hyphae that give a characteristic colour to mature colonies (25). For example, the distinctive blue colour of *Streptomyces coelicolor* colonies comes from a blue pigment related to production of the antibiotic actinorhodin (34). This species is named coelicolor as it is Latin for “heavenly coloured” or “sky coloured” (26).

1.1.3.1. *Streptomyces* development

As well as their complex secondary metabolism, *Streptomyces* bacteria possess a fascinating growth and life cycle, which has also been of particular research interest. Much of the pioneering work on *Streptomyces* genetics was carried out by Sir David Hopwood and co-workers at the John Innes Centre. *Streptomyces* bacteria were originally thought to be fungi, partly because of their similar cell shape and filamentous growth pattern. The morphology of streptomycetes was confirmed through several electron microscopy experiments throughout the 1950s-'60s (35–38). *Streptomyces* cells grow as branching hyphae that form a complex, tightly woven matrix, resulting in vegetative mycelium (25). From mycelial colonies, aerial hyphae form, which act as reproductive structures allowing the dispersal of environmentally resistant spores (30). The spores form as a result of septum formation in the multinucleate aerial hyphae, which

causes separation into individual cells (25). Thus, the *Streptomyces* lifecycle encompasses three developmental stages, from spore to vegetative mycelium to aerial mycelium. The transcription factors controlling the complex regulatory network of sporulation are encoded by the *bld* and *whi* genes (30,39,40). It was recently shown that the secondary messenger metabolite cyclic di-GMP signals through BldD and the sigma factor σ^{WhiG} to control both the formation of reproductive aerial hyphae, and their differentiation into spore chains (41).

Recently, another interesting growth form of *Streptomyces* bacteria was observed, described as the 'explorer' phenotype, which is thought to be triggered by fungal interactions (42). Explorer cells are so-called because of their ability to rapidly spread across both biotic and abiotic surfaces via a non-branching hyphal growth state. Explorer cells can communicate even when physically separated, due to production of the volatile signalling metabolite trimethylamine (42,43). Exploratory growth therefore represents an intriguing method by which *Streptomyces* bacteria might colonise new habitats in the environment.

1.1.3.2. Natural products from *Streptomyces*

As discussed previously, many *Streptomyces*-derived antibiotics were discovered during the golden age of antibiotics, including chloramphenicol from *Streptomyces venezuelae* (44), and tetracyclines such as aureomycin from *Streptomyces aureofaciens* (45). Some important antibiotics for treating multi-drug resistant infections are also produced by streptomycetes, such as daptomycin from *Streptomyces roseosporus* (46,47). *Streptomyces avermitilis* produces avermectins, which are anti-parasitic metabolites used to treat malaria. The development of these drugs led to the award of the 2015 Nobel Prize in Physiology or Medicine to Campbell and Ōmura (48). An anthelmintic metabolite, milbemycin, is produced by *Streptomyces bingchenggensis* (49). Some important antifungal natural products are also produced by streptomycetes, including polyene macrolides such as pimaricin (50) and amphotericin B (51). Other more recently discovered antifungals include the reveromycins from *Streptomyces* sp. 3-10 (52) and the linearmycins produced by *Streptomyces* sp. Mg1 (53). Actinomycin is another important natural product that is produced by *Streptomyces antibioticus*, which became the first bacterial agent used to treat cancer (54). *Streptomyces* bacteria also produce natural products with other important ecological functions, such as the desferrioxamine siderophores that are important for iron-uptake and are also used in medicine to help treat iron deficiency diseases and to help mediate drug delivery (55).

1.1.4. Natural Products from other Kingdoms of Life

1.1.4.1. Fungal natural products

Fungi produce a range of secondary metabolites including maleidrides (56), tropolones (57) and nonribosomal peptides (58,59). As well as producing important antibiotics such as penicillin (13) (**Figure 1**), fungi are also responsible for medicinal metabolites such as the immunosuppressant cyclosporin, produced by *Tolypocladium inflatum* (60). Genera of fungi that have been frequently exploited include *Penicillium*, *Aspergillus* and *Fusarium* (59). However, it is thought that the vast majority of fungal species have not been described or cultured under laboratory conditions (61,62). Much of the early fungal natural product work was carried out by Harold Raistrick (59).

1.1.4.2. Archaeal natural products

Although less well studied, archaea also produce a range of secondary metabolites such as archaeocins and diketopiperazines. Archaea are not known to produce some of the key natural product classes seen in bacteria and fungi such as nonribosomal peptides, but many archaeal natural products show biotechnological potential as antimicrobials, cosmetics and food additives (63).

1.1.4.3. Animal natural products

Animals produce natural products such as defensins, which are cyclic antimicrobial peptides crosslinked by disulphide bonds. These form part of the innate immune system (64). Nematodes such as *Caenorhabditis elegans* were recently shown to produce hybrid polyketide-nonribosomal peptides such as the nemamides (65) (**Figure 1**). Other animal-derived natural products include casein, chitosan and collagen, metabolites with unusual structures and broad applications in industry including catalysis, drug delivery and tissue engineering (66–69).

1.1.4.4. Plant natural products

Like microbes, plants also produce a plethora of natural products, many of which have been utilised in medicine. One of the earliest plant natural products to be developed into a drug was morphine, produced by the 'opium' poppy *Papaver somniferum*, marketed by Merck in 1826 (70,71). Another early and eminent example of a plant natural product is salicin (**Figure 1**), isolated from the tree bark of *Salix alba* (white willow), which was developed into aspirin as a semi-synthetic drug by Bayer in 1899 (70,72). Plants are also particularly prolific producers of terpenes, which have been developed into flavours, scents, cosmetics and medicines (73–75).

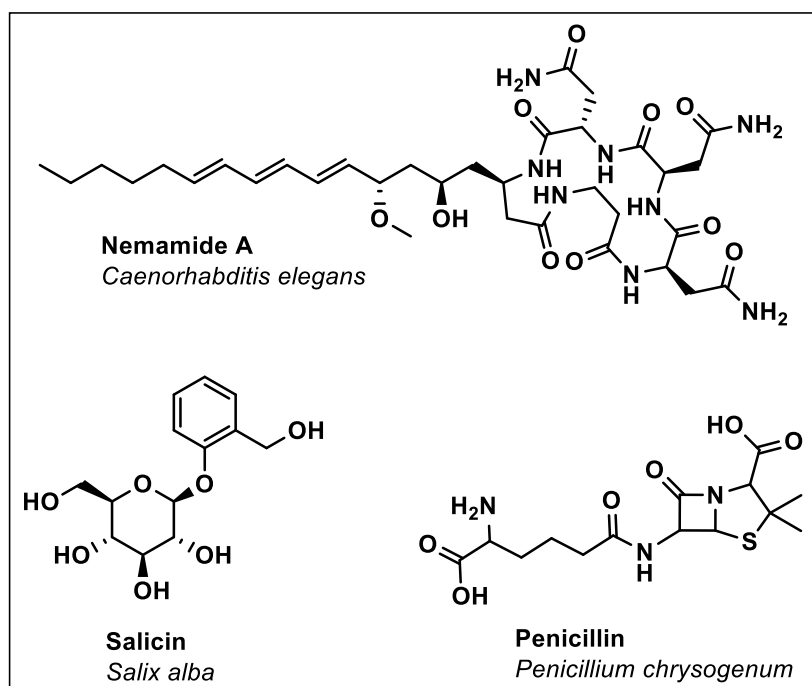


Figure 1: Examples of natural products produced by diverse organisms. Nemamide A is a polyketide-nonribosomal peptide hybrid metabolite produced by *Caenorhabditis elegans*. Salicin is a natural product isolated from *Salix alba* tree bark which is a precursor for the production of aspirin. Penicillin is a β -lactam antibiotic isolated from the fungus *Penicillium chrysogenum*.

1.1.5. Antimicrobial Resistance

A growing problem that is threatening the effectiveness of antimicrobial drugs is antimicrobial resistance. This is the ability of microbes to evolve resistance to antimicrobial drugs through a variety of different mechanisms. Where antibiotics are naturally produced by microbes to kill off their competitors, it follows that microorganisms have also evolved ways to combat this. Fleming famously warned of antibiotic resistance in his 1945 Nobel Lecture: “*It is not difficult to make microbes resistant to penicillin in the laboratory by exposing them to concentrations not sufficient to kill them, and the same thing has occasionally happened in the body.*” This was an early caution for the resistance crisis that is emerging today.

Mechanisms of antimicrobial resistance are diverse, and include modification of a drug target, limitation of uptake, drug inactivation and active efflux from cells. Resistance can also be defined as either intrinsic or acquired (76). Bacteria may be intrinsically resistant to an antibiotic if, for example, they do not naturally harbour the molecular target of the given drug, or because of natural activity of efflux pumps that export the antibiotic. Gram-negative bacteria are also intrinsically resistant to many antibiotics due to the thick lipopolysaccharide layer that reduces permeability of the outer membrane (77,78). Acquired resistance refers to the gaining of genetic material through horizontal gene

transfer that confers resistance (76), but it can also arise spontaneously. The spread of acquired resistance is accelerated through increased exposure to antimicrobials. For this reason, the over-prescribing and misuse of antibiotics in clinical and agricultural settings is largely attributed to the rising prevalence of antimicrobial resistance (79). However, evidence also suggests that resistance even predates human usage of antibiotics. For example, antibiotic resistance elements have been found in ancient DNA dating from the Pleistocene epoch (80), in deep sea environments (81) and in a cave that had been secluded for 4 million years prior to investigation (82).

As laid out in the 2016 O'Neill report, it is often reported that by 2050 there will be more than 10 million deaths per year from antimicrobial resistance (83). But how predictable is the future? There are many unpredictable factors that could rapidly change the resistance landscape, which are described by David Livermore as 'Black Swan' events. These include the escape of new resistance genes from environmental bacteria and their uptake by promiscuous mobile genetic elements (84). An example of this is the escape of the *vanHAXY* operon, putatively from *Paenibacillus* spp. to Tn1546, resulting in the unexpected emergence of vancomycin-resistant Enterococci (VRE) (85), which now represent major clinical pathogens. The unpredictable nature of antimicrobial resistance makes it an even more pressing issue, and it is vital that we are equipped to deal with future challenges.

1.1.6. Approaches to Natural Products Discovery

Whilst tackling antimicrobial resistance requires a multifaceted approach, one essential element is the discovery of new antimicrobial agents. Antibiotic discovery comes with many practical, regulatory and economical challenges, and as a result the pharmaceutical industry has largely cut investment in antibiotic and natural product discovery programs (86). However, there is still a huge amount of potential in natural products from microorganisms. Recently, with the inception of the genomic era, it has been revealed through analysis of microbial genomes that their biosynthetic capacity has not been fully explored, and that they are capable of producing many more natural products than are currently characterised (87,88). For this reason, natural products research remains a promising arena in which to discover novel antimicrobials and other useful natural products, while also developing a deeper understanding of their biosynthesis and ecological functions. A common problem with traditional activity-based screening for antimicrobials is the high rate of rediscovery of already known metabolites. Waksman wrote about this duplication problem in 1965 (89). This has led to research groups taking alternative approaches to natural products discovery in recent years. One way to discover new natural products is through genome mining for novel biosynthetic

gene clusters, which is discussed in detail in Chapter 2. Other approaches are described in the following sections.

1.1.6.1. Exploring new environments

Some natural product research has focused on studying diverse and more extreme environments, such as desert (90), arctic (91) and marine (92) habitats, in order to isolate novel microorganisms and investigate their biosynthetic capacity. Other research has also focused on investigating interesting ecological niches where microorganisms co-exist in a symbiotic relationship with other species. One example is that of the leafcutter ants, which harbour unique *Streptomyces* species on their surfaces. When isolated, it was found that these bacteria produce several types of antibiotic metabolites including the fasamycins and formicamycins (93). The genomes of the newly identified *Streptomyces* species were also analysed indicating that they were taxonomically distinct, and contain many biosynthetic gene clusters for natural products that were not seen in other species (94). This suggests that investigating bacteria from unique habitats is a promising way to identify structurally novel metabolites. This is also exemplified with the recent discovery of the antibiotic darobactin, which was isolated from *Photorhabdus khanii* HGB1456. *Photorhabdus* is an underexplored genus that live symbiotically within entomopathogenic nematodes (95).

1.1.6.2. Novel isolation methods

Whilst the identification of novel bacterial species is an exciting venture, one problem that thwarts microbial natural products discovery is that a large proportion of microbes are difficult to grow and are deemed uncultivable under lab conditions: the so-called 'great plate count anomaly' (96). This can make subsequent isolation and investigation of any natural products challenging. However, some novel approaches have been taken to try to overcome the cultivation problem. One of these is the development of the isolation chip (iChip), a device containing wells that is placed in a natural environment such as the soil, allowing microorganisms to directly grow in the presence of natural nutrients and growth factors (97). This approach was developed by Kim Lewis's lab leading to the isolation of a previously undiscovered bacterium *Eleftheria terrae*, which produced the novel antibiotic teixobactin. Teixobactin has activity against some Gram-positive pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA), and is thought to be unlikely to induce resistance as it adopts a unique interaction with the precursor of bacterial peptidoglycan (98).

1.1.6.3. Direct analysis of environmental samples

Another approach to overcome cultivation problems is the direct capture of metagenomic DNA from environments such as the soil, in order to assemble genomes, identify biosynthetic gene clusters and express them for further characterisation. This approach has been taken by the Brady lab who identified the fasamycin antibiotics (99). In some cases environmental samples have also been directly analysed by mass spectrometry (92), and natural products were detected from these with the help of public mass spectral databases such as GNPS (Global Natural Products Social) molecular networking (100) (discussed in further detail in Chapter 2).

1.1.7. Natural Product Classes

Natural products are grouped into different subclasses based on their biosynthetic and structural features and can be broadly divided into nonpeptides and peptides. Nonpeptide natural products include polyketides and terpenes, and peptide natural products include nonribosomal peptides and ribosomally synthesised peptides.

1.1.7.1. Polyketides

Polyketides are a major natural product class comprising metabolites such as polyethers, polyenes, polyphenols, macrolides and enediynes. They have diverse bioactivities including antimicrobial, anticancer, immunosuppressive, anti-cholesterol and anti-inflammatory. Polyketides are produced by a range of organisms including bacteria, fungi, plants, protists, insects molluscs and sponges (101–103). The biosynthesis of polyketides is similar to that of fatty acids and follows assembly-line biosynthesis. Assembly-lines consist of complexes of large proteins comprising repetitive peptide sequences, each of which encode individual enzyme domains. There are several different architectures of the polyketide synthase (PKS) assembly line: type I PKSs (T1PKS) e.g. erythromycin (**Figure 2**) (104), type II PKSs (T2PKS) e.g. actinorhodin and type III PKSs (T3PKS) e.g. plant chalcone synthase. PKSs can also be categorised as iterative or noniterative, depending on whether each ketosynthase domain catalyses more than one round of elongation during biosynthesis (102). PKSs typically consist of a ketosynthase (KS)-domain, an acyltransferase (AT)-domain and an acyl carrier protein (ACP). In T1PKSs, the AT-domain uptakes acyl-coenzyme A (CoA) substrates, which are then transferred to the phosphopantetheine of the ACP. The KS-domain is then transacylated with the growing polyketide chain, enabling malonyl ACP Claisen condensation with the acyl KS-substrate. Each PKS module thereby incorporates a C2 acetyl unit, and the newly formed β -keto group usually undergoes reduction from a ketoreductase, dehydratase or enoyl reductase (105,106). In T2PKSs, acetate is incorporated into the ACP and then transferred to the active site of a KS, which then

undergoes iterative elongation by Claisen condensation with derivatives of malonyl CoA (107). The polyketide is then cyclised to produce polycyclic metabolites (102).

1.1.7.2. Terpenes

Terpenes, or isoprenoids, are the largest and most diverse class of small metabolite natural products, produced by almost all kingdoms of life, particularly plants. They carry out a range of biological functions in defence, photosynthesis and electron transfer (73,108). Terpenes originate from the C₅ substrates dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP). DMAPP is usually condensed with one or more IPP metabolites in a head-to-tail manner to form geranyl diphosphate (GPP) (C₁₀), farnesyl diphosphate (FPP) (C₁₅) or geranylgeranyl diphosphate (GGPP) (C₂₀). FPP and GGPP can then condense in a head-to-head manner to form dehydrosqualene, squalene or phytoene. These serve as the precursors for carotenoids, sterols and hopanoids. Isoprenoids can also be cyclised to produce terpene natural products such as monoterpenes, diterpenes and sesquiterpenes (108). Plants can also convert DMAPP into isoprene, a potential source of renewable fuel (73,109). DMAPP and IPP can be made via two different pathways: the mevalonate pathway, utilised by most eukaryotes and Archaea, or the methylerythritol phosphate pathway, utilised by most bacteria (108). Plants have been found to utilise both pathways (110). Enzymes involved in terpene biosynthesis include prenyl transferases, terpene synthases and 4Fe-4S reductases (109). In plants, it has also been found that non-terpene synthase enzymes can catalyse the formation of terpenes (111). An example of a microbial terpene is geosmin (**Figure 2**), produced by cyanobacteria and actinomycetes from the sesquiterpene precursor farnesyl pyrophosphate (112).

1.1.7.3. Nonribosomal peptides

Nonribosomal peptides are peptide natural products produced by nonribosomal peptide synthetases (NRPSs), which are mainly found in bacteria and fungi. Nonribosomal peptides include a range of structurally complex metabolites, such as head-to-tail cyclised peptides, lipocyclopeptides and linear peptides of varying size. Nonribosomal peptides can also undergo extensive structural modifications by tailoring enzymes, producing many antibiotic nonribosomal peptides including the β -lactams and glycopeptides such as vancomycin (**Figure 2**) (113). Nonribosomal peptides also display other biological functions, acting as siderophores, toxins, antifungal and antitumour agents (114). Nonribosomal peptide biosynthesis involves three main stages: building block assembly, NRPS-mediated peptide assembly and post-NRPS modification and decoration. NRPSs are multi-modular enzymes with each typical module including an adenylation (A)-domain, condensation (C)-domain and thiolation (T)- or peptidyl carrier

protein (PCP)-domain. The A domain activates the building block by adenylation, and the activated building block is transferred to the pantetheine thiol group of the PCP. The C domain transfers an aminoacyl or peptidyl group from an upstream carrier protein domain to the primary amine of an amino acid that has been previously loaded onto the downstream carrier domain. The thioesterase domain catalyses either hydrolysis or cyclisation of the peptide to release it from the final carrier protein domain (114,115).

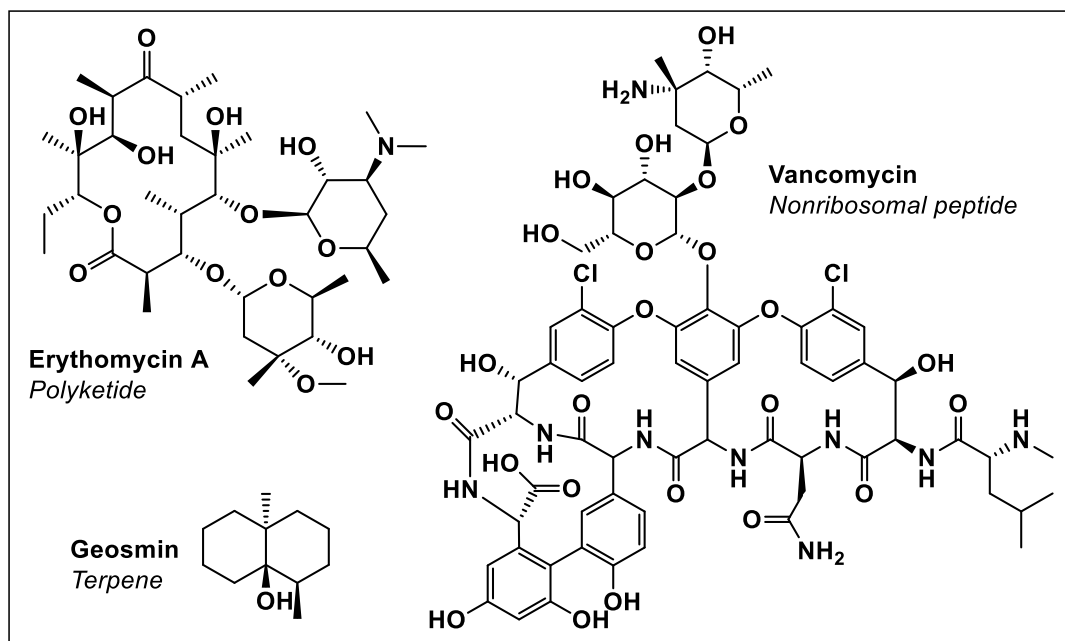


Figure 2: Examples of metabolites from different natural product classes. Erythromycin A is a type 1 polyketide antibiotic produced by *Streptomyces erythreus* (104), vancomycin is a nonribosomal glycopeptide antibiotic produced by *Nocardia orientalis* (113) and geosmin is an irregular sesquiterpene that has a characteristic earthy odour, produced by actinomycetes and cyanobacteria (112).

1.1.7.4. Ribosomally synthesised and post-translationally modified peptides (RiPPs)

Other peptide natural products include the ribosomally-synthesised and post-translationally modified peptides (RiPPs). RiPPs are the focus of this thesis, and will therefore be described in detail in the following sections.

1.2. RiPPs

1.2.1. General Biosynthetic Logic

RiPPs are a class of peptide natural product that are synthesised on the ribosome and subject to extensive post-translational modifications, giving rise to a high degree of structural complexity and biological diversity. RiPPs are produced from a short precursor peptide, which is typically 50-150 amino acids in length. This precursor usually comprises a leader region and a core region. The core region includes the amino acid residues that are ultimately post-translationally modified by a series of RiPP tailoring enzymes, which decorate the peptide backbone with various structural features (116,117). Once the core region has been fully modified, it is cleaved from the leader region, yielding the biologically active final natural product (116,117) (**Figure 3**).

The leader region of the precursor peptide has been suggested to play multiple roles in RiPP biosynthesis. Firstly, it has been shown to be important for the binding of RiPP tailoring enzymes, which contain domains called RiPP precursor peptide recognition elements. These recognition elements are thought to be specific to certain classes of biosynthetic enzymes (118–120). The leader peptide has also been suggested to be involved in protection, preventing the core region from proteolytic cleavage before the biosynthetic modifications are complete. In some rare cases, RiPP precursors contain a follower region instead of a leader region, as is seen in bottromycins (121). In other cases, precursors contain both a follower and leader region, as is the case for pantocin A (122), α -amanitin (123), thiovarsolins (124) and some cyanobactins (125). As with other microbial natural products, the biosynthetic genes for RiPPs typically cluster together tightly on the genome, and the precursor peptide is generally denoted as 'xxxA'. Due to the diversity of RiPP metabolites, they are grouped into various different subclasses based on their biosynthetic and structural features.

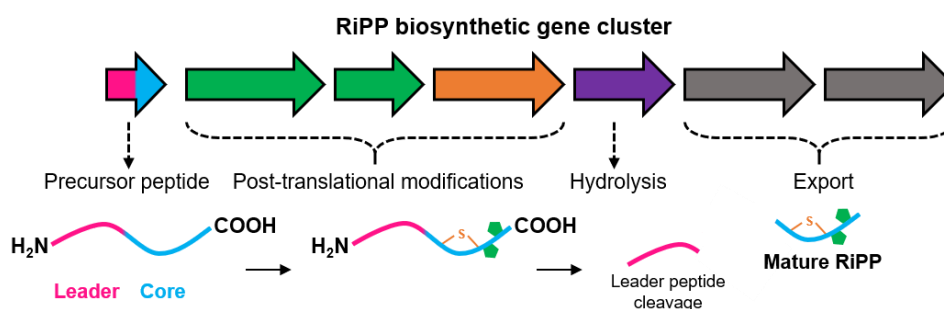


Figure 3: General biosynthetic logic of ribosomally synthesised and post-translationally modified peptides (RiPPs). A precursor peptide comprising a core peptide and leader and/or follower peptide is subject to post-translational modifications in the core region by a series of RiPP tailoring enzymes. The leader/follower peptide is then cleaved from the core peptide to yield the mature modified peptide.

1.2.2. RiPP Subclasses

1.2.2.1. Linear azol(in)e-containing peptides (LAPs)

Linear azol(in)e-containing peptides (LAPs) are a class of RiPP characterised by the presence of azole or azoline heterocycles on a linear peptide backbone. The first LAP discovered was the β -haemolytic factor streptolysin S, produced by *Streptococcus pyogenes*, which was first observed in 1901. The defining genetic and structural features were not reported until 2000 (126). Another well-studied LAP is microcin B17 (**Figure 4**), a bactericidal peptide produced by *Escherichia coli* which was first observed in 1976 (127), and displays narrow spectrum antibacterial activity targeting DNA gyrase (128). Microcin B17 is a large metabolite deriving from a 43-residue core amino acid sequence, which is modified to contain four oxazole and four thiazole heterocycles. A more recently discovered example of a LAP is klebsazolicin, which was identified following genome mining for homologues of the microcin B17 biosynthetic proteins (129). Klebsazolicin was isolated from *Klebsiella pneumoniae* sp. *ozaenae* in 2017 and displays antibiotic activity via ribosome inhibition. Klebsazolicin derives from a shorter 23-residue core amino acid sequence, which is modified to contain one oxazole and three thiazole heterocycles, as well as an additional N-terminal amidine ring (129).

1.2.2.2. Thiopeptides

Thiopeptides, also known as thiazolylpeptides, are characterised by a macrocyclic core comprising a central six-membered nitrogen-containing heterocycle and a series of thiazoles. Further biosynthetic features include the dehydration and cyclodehydration of serine, threonine and cysteine residues (116). Thiopeptides have a range of biological activities such as antibacterial and anticancer (130,131), antiplasmodial (132) and immunosuppressive (133). One of the first thiopeptides identified was micrococцин (134), which was isolated in 1948 from sewage water (135), and is produced by *Bacillus* and *Micrococcus* species (136). Micrococцин is an antibiotic which also displays inhibitory activity against the malaria parasite *Plasmodium falciparum* (137). Another early example of a thiopeptide is thiostrepton, isolated from *Streptomyces azureus* in 1954 (138) and produced by several other *Streptomyces* species (139). The structure of thiostrepton (**Figure 4**) was elucidated by Dorothy Crowfoot Hodgkin in 1970 using X-ray crystallography (140). Other examples of thiopeptides include the thiocillins, produced by *Bacillus cereus* (141).

1.2.2.3. Cyanobactins

Cyanobactins are RiPPs produced by cyanobacteria, whose structural transformations often include heterocyclisation, oxidation or prenylation of amino acid residues. The most well-studied cyanobactins are the patellamides (**Figure 4**) and trunkamides, which

display cytotoxic activities. The patellamides are produced by *Prochloron* species, (142) and trunkamide was isolated from *Lissoclinum patella* in 1996 (143). Patellamides contain pseudosymmetrical cyclic dimers with thiazole and oxazoline heterocycles installed in between nonpolar amino acids. Trunkamides often contain proline, thiazolines and prenylated serine and threonine derivatives (142). While many cyanobactins are cyclic peptides, more recently several linear cyanobactins have also been reported. Scytodecamide is a linear decapeptide cyanobactin produced by the freshwater cyanobacterium *Scytonema* sp. UIC 10036, which was reported in 2019 (144). Scytodecamide contains an N-terminal N-methylation and a C-terminal amidation (144). As well as cyanobactins, cyanobacteria also produce a class of RiPP called microviridins, which contain lactone and lactam groups (116).

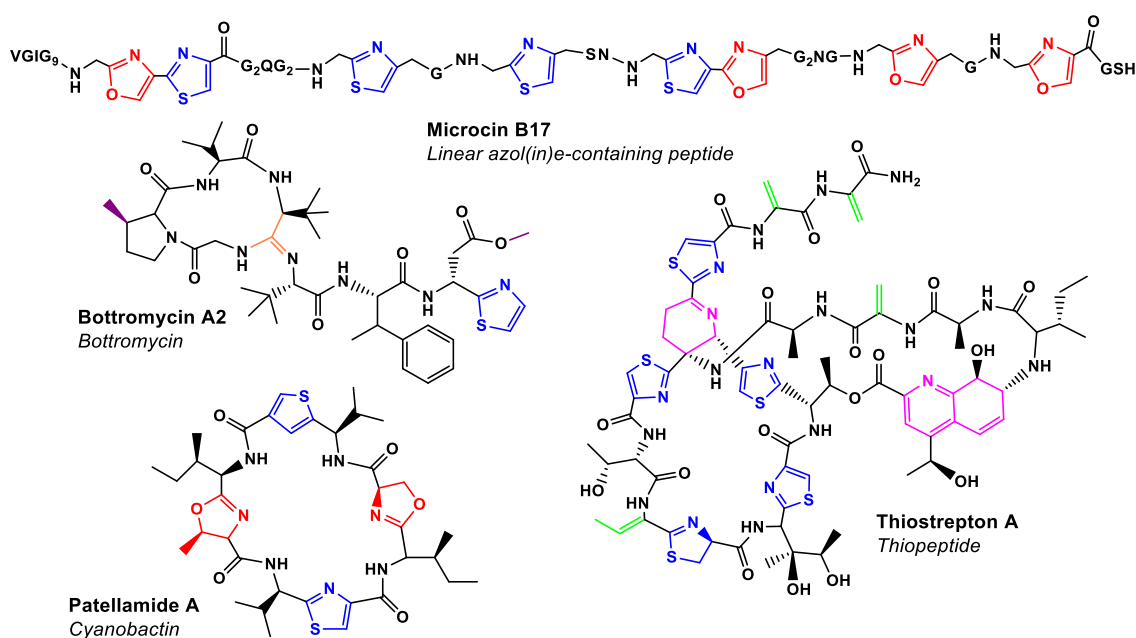


Figure 4: Chemical structures of RiPPs containing (methyl)oxazole and thiazole heterocycles. Microcin B17 is an antibiotic linear azol(in)e-containing peptide (LAP) produced by *Escherichia coli*, bottromycin A2 is an antibiotic produced by *Streptomyces* species, thiostrepton A is a thiopeptide antibiotic produced by *Streptomyces* species and patellamide A is a cyanobactin produced by *Prochloron* species. Oxazoles are highlighted in red and thiazoles are highlighted in blue. Other highlighted post-translational modifications include methylation (purple) and amidine ring formation (orange) in bottromycin, and dehydration (green) and nitrogen-containing heterocycle formation (pink) in thiostrepton A.

1.2.2.4. Bottromycins

Bottromycins, exemplified by bottromycin A2 (**Figure 4**), are a small subclass of RiPP metabolites that have been of great research interest due to their unique structure and antibiotic activity. Bottromycin was first isolated from *Streptomyces bottropensis* in 1967 (145) and is produced by other *Streptomyces* species such as *S. scabies* and

Streptomyces sp. WMMB272 (146). Bottromycin exerts antibiotic activity via ribosome inhibition and has been shown to be effective against MRSA and VRE. The structure of bottromycin includes a unique N-terminal macroamidine ring, as well as a C-terminal thiazole heterocycle. Derivatives of bottromycin contain differing methylation patterns: bottromycin A2 contains a single C-methylation on proline, bottromycin B lacks any methylation on proline and bottromycin C contains a doubly methylated proline (147).

1.2.2.5. Thioamitides

Thioviridamide is a thioamidated RiPP produced by *Streptomyces olivoviridis* NA005001. Its unique structure includes structurally rare thioamide bonds and two unusual amino acids: β -hydroxy-N1,N3-dimethylhistidinium (hdmHis) and S-[(Z)-2-aminovinyl]-D-cysteine (AviCys) (148). Thioviridamide was first isolated in 2006 (148), but its gene cluster was not reported until 2013 (149). Thioviridamide possesses potent antiproliferative activity against cancer cells, which is attributed to its unusual structure (148). Recent genome mining approaches have unveiled several further thioviridamide-like metabolites, which have recently been termed thioamitides (150). These include thioalbamide from *Amycolatopsis alba* (151) (**Figure 5**) and the thiolgamides from *Streptomyces malaysiense* (152). These recently identified thioamitides also appear to exhibit cytotoxic activity against cancer cell lines (151,152).

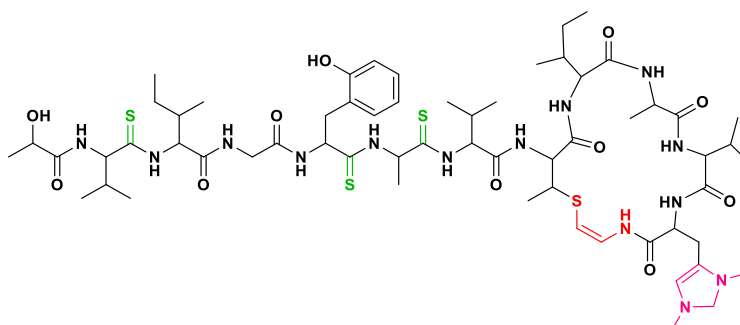


Figure 5: Chemical structure of thioalbamide, a recently reported thioamitide produced by *Amycolatopsis alba*. The characteristic thioamide bonds are highlighted in green. The structure also features the unusual amino acids S-[(Z)-2-aminovinyl]-D-cysteine (AviCys, red) and β -hydroxy-N1,N3-dimethylhistidinium (hdmHis, pink).

1.2.2.6. Lanthipeptides

Lanthipeptides are a large and well-studied group of RiPPs characterised by the presence of sulphur-to- β -carbon thioether cross-links named lanthionines (Lan) and methyllanthionines (MeLan). Lanthipeptides are divided into five classes (I-V), based on the biosynthetic machinery responsible for installing the thioether rings. Lanthipeptides display diverse activities, and those with antibiotic properties are termed lantibiotics. A

characteristic lanthipeptide is the polycyclic antibiotic peptide nisin (**Figure 6**). Its structure derives from a 34-residue core peptide, which is post-translationally modified to contain Lan and MeLan as well as the unusual amino acids didehydroalanine (Dha), didehydroaminobutyric acid (Dhb) and α -aminobutyric acid (Abu) (153). A more recently reported lantibiotic is kyamicin, which was isolated following heterologous expression of a cryptic biosynthetic gene cluster from a plant-ant derived *Saccharopolyspora* species. Kyamicin displays antibiotic activity against *Bacillus subtilis* EC1524 (154). Whilst lanthipeptides have been widely studied for many years, the class V lanthipeptides were only recently defined in 2020, with the discovery of a novel metabolite with unique biosynthetic machinery from *Streptomyces pristinaespiralis* ATCC 25468 (155).

1.2.2.7. Sactipeptides

Another class of RiPP that harbours thioether linkages are the sactipeptides. In this class the thioether bonds comprise sulphur-to- α -carbon cross-links. This was first reported in 2003 with the structural elucidation of subtilisin A, produced by *B. subtilis* (156). Since then, many other sactipeptides have been discovered such as thurincin H from *Bacillus thuringiensis* (157) and ruminococcin C, an anti-clostridial metabolite produced by *Ruminococcus gnavus*, a bacterium found in the human microbiome (158). Sactipeptides with antibiotic activity are termed sactibiotics (116).

1.2.2.8. Linaridins

Linaridins are linear dehydrated peptides, which are a small but growing subclass of RiPP. The archetypal member of this class is cypemycin (**Figure 6**), isolated from *Streptomyces* sp. OH-4156 in 1993 (159). Cypemycin displays antibiotic activity and was originally considered a lantibiotic due to the presence of Dhb and AviCys residues in its structure. However, further genetic studies showed that these modifications are carried out via unusual biosynthetic steps unrelated to lanthipeptide biosynthesis (160). Cypemycin was thus described as the founding member of the new linaridin class. As well as Dhb and AviCys, cypemycin also contains unusual N,N-dimethylalanine and allo-isoleucine residues installed onto the 22 amino acid core peptide (161). Following cypemycin discovery, further homologues were reported, including grisemycin from *Streptomyces griseus* (162) and legonaridin from *Streptomyces* sp. CT34 (163). More recently in 2019, the salinipeptides were discovered from *Streptomyces* sp. GSL-6C, a strain isolated from the Great Salt Lake (164).

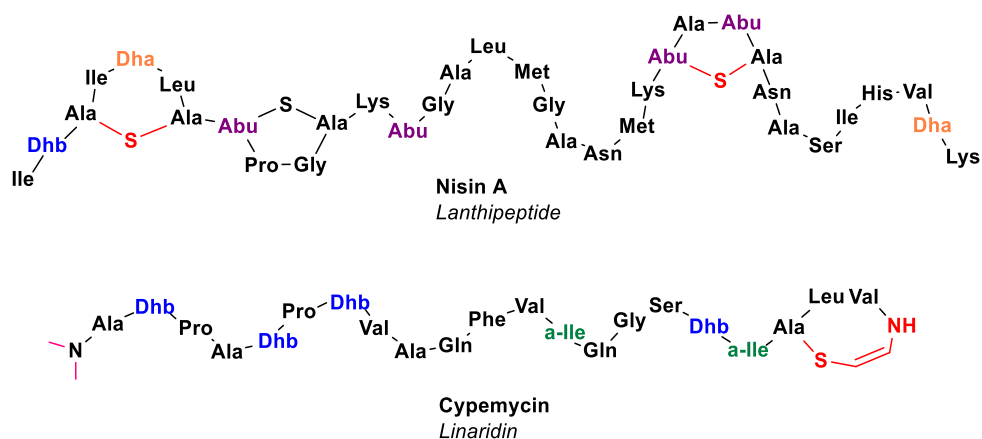


Figure 6: Chemical structures of nisin A and cypemycin. Nisin A is a lanthipeptide with antibiotic activity produced by *Streptococcus lactis*, and cypemycin is a linaridin with antibiotic activity produced by *Streptomyces* sp. OH-4156. Less common amino acids in these metabolites are highlighted, including dihydroalanine (Dha, orange), dihydroaminobutyric acid (Dhb, blue), aminobutyric acid (Abu, purple) and allo-isoleucine (green). Other post-translational modifications include (methyl)lanthionine bonds in nisin A (red) and an S-[(Z)-2-aminovinyl]-D-cysteine (AviCys) moiety in cypemycin (red). Cypemycin also contains an unusual N,N-dimethylalanine moiety (pink).

1.2.2.9. Lanthidins

Recently in 2020, a new subclass of RiPP was described with the identification of a novel metabolite cacaoidin, which harbours elements of both lanthipeptides and linaridins. Cacaoidin features several D-amino acids, an unusual glycosylated tyrosine residue and a unique N,N-dimethylanthionine residue. Cacaoidin also displays antimicrobial activity against clinical pathogens such as *Clostridium difficile*. Although this metabolite contains elements of lanthipeptides and linaridins, the biosynthetic gene cluster of cacaoidin shows low homology to these classes, suggesting that lanthidins are a unique new family of RiPP (165).

1.2.2.10. Lasso peptides

Lasso peptides are cyclic metabolites that display antibiotic activities and contain a characteristic knotted 'lasso' structure. This is made up of a right-handed N-terminal macrolactam ring comprising seven to nine amino acid residues, through which a linear C-terminal peptide tail is threaded through in a noncovalent interaction (166). There are three different subclasses of lasso peptide based on the number of disulphide bonds in the metabolite. Class I lasso peptides contain two disulphide bonds, class III contain one disulphide bond and class II have no disulphide bonds (167). A model lasso peptide is microcin J25 (**Figure 7**), an RNA inhibitor identified from *E. coli* in 1992, which is active against Gram-negative bacteria (168). Microcin J25 derives from a 21-residue peptide, but the exact structural confirmation of the mature metabolite was contested for some

years (169), initially thought to comprise a head-to-tail cyclised backbone, before two research groups proved the lasso tail structure in 2003 (166,170). Lasso peptides discovered more recently in 2019 include ubonodin, which has antimicrobial activity against members of the *Burkholderia cepacia* complex (171), and pandonodin, a lasso peptide with an unusually long 18-residue C-terminal tail (172).

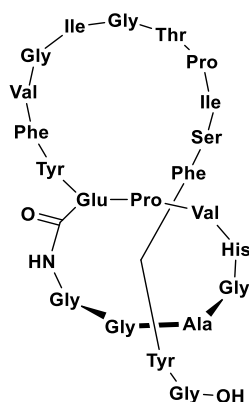


Figure 7: Chemical structure of microcin J25, a characteristic lasso peptide produced by *E. coli*. The structure includes a macrolactam ring through which the C-terminal tail is threaded.

1.2.2.11. Streptide and darobactin

Streptide is a macrocyclic peptide that contains an unusual lysine-to-tryptophan carbon-carbon crosslink (**Figure 8**). Streptide was first reported in 2007 as a modified nine-residue peptide termed Pep1357, produced from *Streptococcus thermophilus* (173). The peptide is excreted as a pheromone and its transcription is controlled via quorum sensing (173). In 2015, detailed biosynthetic and structural elucidation of streptide was finally reported, revealing the novel crosslinking between lysine and tryptophan side chains (174) and further stereochemistry was confirmed in 2019 (175). Until recently, streptide was a structurally unique RiPP, but in 2019 a novel metabolite called darobactin was reported (**Figure 8**), which also features lysine-to-tryptophan crosslinking. In addition, darobactin also harbours an unprecedented aromatic–aliphatic ether link between two tryptophan residues in the core peptide. Darobactin displays antibiotic activity against Gram-negative pathogens including *E. coli*, *Klebsiella* and *Pseudomonas* species (95). This discovery thus highlights expansion of this structurally rare RiPP class.

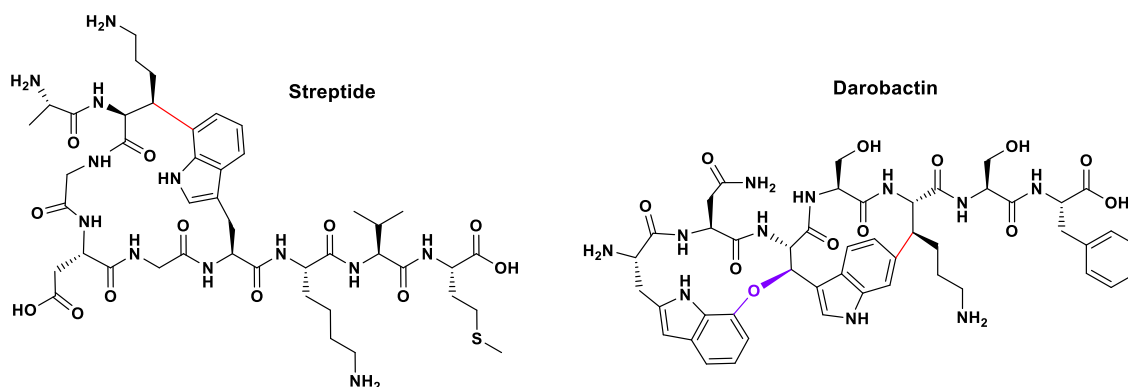


Figure 8: Chemical structures of streptide and darobactin. Streptide is a pheromone produced by *Streptococcus thermophilus* and darobactin is an antibiotic produced by *Photobacterium khanii*. The tryptophan-lysine crosslinking unique to these metabolites is highlighted in red, and the unusual ether bond in darobactin is highlighted in purple.

1.2.2.12. Proteusins

Proteusins are exemplified by the polytheonamides, first reported in 1994 (176). These metabolites are produced by the Japanese sponge *Theonella swinhoei*, which harbours a range of symbiotic bacteria. Polytheonamides are large and highly complex metabolites, containing several nonproteinogenic and D-configured amino acids and an unusual N-acyl moiety. Other post-translational modifications include dehydration, methylation and hydroxylation (176–178). Polytheonamides are highly cytotoxic owing to their ability to form membrane channels by adoption of a β -helical secondary structure (179).

1.2.2.13. Other bacterial RiPPs

As well as the diverse subfamilies described thus far, bacteria produce many more classes of RiPP. These include the relatively large bacterial head-to-tail cyclised peptides, and much smaller metabolites such as the bacterial cofactor pyrroloquinoline quinone (PQQ) (180), the pantocins (122) and thyroid hormones such as 3,30,5,50-tetraiodothyronine (T4) (116). Whilst the majority of these RiPPs are produced by bacteria, further RiPP subfamilies have also been described from other organisms.

1.2.2.14. Fungal RiPPs

In fungi, four different families of RiPP have been described. The amatoxins and phallotoxins were first reported in 2007, and are produced by basidiomycete genera such as *Amanita*, *Glaerina*, *Lepiota* and *Conocybe* (181–183). Amatoxins, exemplified by α -amanitin, are highly toxic to insects, nematodes and mammals, acting via RNA inhibition (183). Amatoxins and phallotoxins consist of eight or seven amino acid residues respectively, forming a bicyclic structure due to crosslinking between cysteine and

tryptophan residues (123). This crosslink is called a tryptathionine (184). Other post-translational modifications include hydroxylations and epimerisation (183). In 2016 the dikaritins were defined, which are cyclopeptides produced by fungi of the Dikarya subkingdom. These RiPPs are cyclised via ether bridges between the hydroxyl group of tyrosine and the β -carbon of isoleucine, phenylalanine or tyrosine (183,185). Another fungal RiPP family are the borosins, recently described in 2017. The founding member, Omphalotin A, is a toxic cyclic dodecapeptide produced by the basidiomycete *Omphalotus olearius*. Nine of the 12 backbone residues in this metabolite undergo N-methylation (186–188). Another fungal RiPP family are the epichloëcyclins, which are cyclic nonapeptides produced by *Epichloë* ascomycetes. These metabolites are cyclised between a conserved tyrosine and proline or isoleucine at the aminoterminal. Dimethylation occurs on a conserved lysine (183,189).

1.2.2.15. Plant RiPPs

In plants, several cyclic RiPP families have been described. Cyclotides are head-to-tail cyclised peptides with three disulphide bonds, forming a characteristic cyclic cysteine knot structure. Cyclotides have been isolated from *Viola* and *Oldenlandia* species (190). PawS-derived peptides are cyclotides with one disulphide bond, which are widespread in the Asteraceae family (191). Orbitides are another subclass of head-to-tail cyclised peptide that do not contain disulphide bonds, and are produced by *Linum usitatissimum* L. (flaxseed) (192). A further plant RiPP family, the lyciumins, has also recently been defined. Lyciumins feature a distinctive N-terminal pyroglutamate and a rare macrocyclic linkage between a C-terminal nitrogen from tryptophanindole and a glycine α -carbon (193). Lyciumins are produced by plants such as *Lycium barbarum* and *Ceolsia argentea*, and are used as a Chinese medicine to treat hypertension, acting via inhibition of angiotension-converting enzyme and renin (194).

1.2.3. Biosynthesis of Post-Translational Modifications in RiPPs

1.2.3.1. *Understanding RiPP biosynthesis*

The focus of this thesis is to identify and characterise novel RiPP biosynthetic gene clusters. In order to mine for gene clusters that might produce structurally novel metabolites, it is important to understand the classes of enzyme that produce the wide-ranging post-translational modifications of RiPPs. For example, how widespread are the biosynthetic genes in nature? How promiscuous or diverse is the enzyme's catalytic activity? Is there potential to expand on a structural class that is currently underexplored? The biosynthetic machinery responsible for many common RiPP post-translational modifications are described in the following sections.

1.2.3.2. *Oxazol(in)e and thiazol(in)e rings*

Azole and azoline heterocycles are characteristic features of RiPPs such as LAPs, thiopeptides and cyanobactins. (Methyl)oxazolines are oxygen-containing rings derived from the side chains of serine and threonine residues, while thiazolines are sulphur-containing rings derived from the side chain of cysteine. These rings can then be oxidised to the respective (thi/ox)azole heterocycle. Much of the understanding of azol(in)e biosynthesis comes from the study of the LAP microcin B17. Azoline ring formation is catalysed by a YcaO-domain protein in collaboration with a partner protein, which is often annotated in genomes as an E1-ubiquitin activating enzyme or an Ocin-ThiF-like protein (195). The YcaO-domain protein is responsible for enzymatic cyclodehydration, while the partner protein is a docking element thought to be involved in peptide recognition. YcaO proteins use ATP as a substrate, phosphorylating the amide backbone via a hemiorthoamide intermediate before subsequent ring formation and elimination of the phosphate (**Figure 9A**) (196–198). While many YcaO-domain proteins are standalone enzymes, in some cases the YcaO and E1 proteins function as a fused cyclodehydratase. This is seen in the cyanobactin enzymes TruD and PatD, which catalyse the biosynthesis of trunkamides and patellamides respectively (75,86). (Methyl)oxazolines and thiazolines can be oxidised by a flavin mononucleotide (FMN)-dependent dehydrogenase to the respective (methyl)oxazoles and thiazoles. The dehydrogenase, E1-like protein and YcaO-domain protein are sometimes described as a TOMM (thiazole/oxazole-modified microcin) synthetase (200) or a BCD complex: the dehydrogenase (B) protein, E1-like (C) protein and YcaO-domain (D) protein were named based on the order of genes in the microcin B17 gene cluster, alongside the precursor peptide (A) (116,201,202).

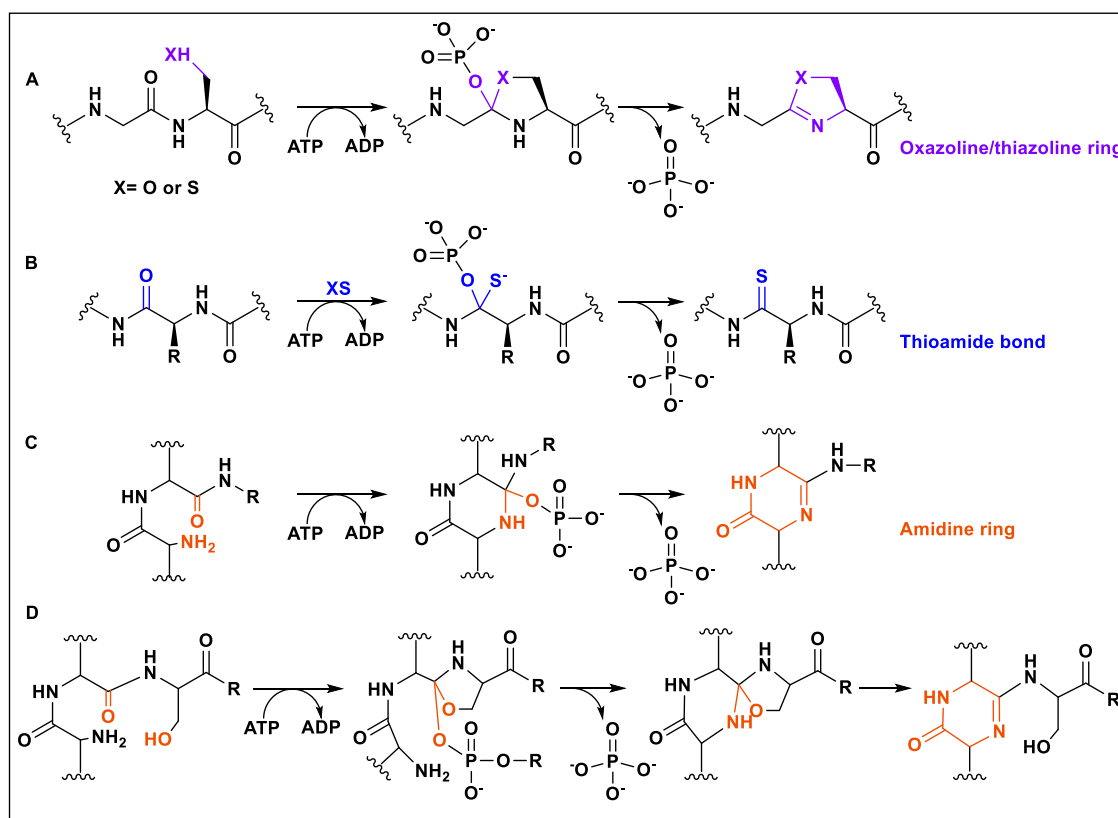


Figure 9: Mechanisms of YcaO-domain protein-mediated catalysis. A: oxazoline (X=O) and thiazoline (X=S) ring formation (197), B: thioamide bond formation (203), C: amidine ring formation via a one-step mechanism (204) and D: amidine ring formation via a possible two-step mechanism (205).

1.2.3.3. Thioamide bonds

Thioamidation is a rare post-translational modification of RiPPs, where the oxygen atom in an amide bond is replaced with sulphur. One example of thioamidation is known outside of natural products, in the archaeal protein methyl-coenzyme M reductase (206). In thioviridamide and other thioamitidies, thioamidation is catalysed by a YcaO-domain protein with a TfuA partner protein (124,197,207). It is thought that thioamidation occurs through a similar mechanism to azoline formation, via an O-phosphorylated hemiorthoamide intermediate (203). Subsequently, an exogenous source of sulphide acts as a second nucleophile, attacking the activated carbonyl to form a tetrahedral intermediate, followed by elimination of the phosphate (**Figure 9B**) (196–198). The TfuA partner protein might act to facilitate binding to the peptide substrate, or to help deliver sulphide units (208). Interestingly, some RiPPs contain thioamide moieties but lack a YcaO-domain protein in the biosynthetic pathway, such as for the chalkophore methanobactin. This suggests that alternative pathways exist for the formation of thioamides. The methanobactin biosynthetic gene cluster consists of a *mbnABC* operon, encoding the precursor peptide (MbnA) and a heterodimer comprising a DUF692 family iron enzyme (MbnB) and a protein from a previously unknown family (MbnC). It was

shown that the oxazolone-thioamide moiety in methanobactin was introduced through dioxygen-dependent four-electron oxidation of the precursor peptide via a metalloenzyme-mediated radical mechanism (203,209).

1.2.3.4. Amidine rings

Amidine rings are another rare post-translational modification found in two known RiPPs, bottromycin (147) and klebsazolicin (129). Amidines are nitrogen-containing heterocycles derived from the dehydration reaction between amine and carbonyl groups of amino acids in the core peptide. In the two known RiPP examples, amidine formation is also biosynthesised by a YcaO-domain protein. In bottromycin biosynthesis, one of two divergent standalone YcaO-domain proteins in the biosynthetic gene cluster is responsible for macroamidine formation, while a different YcaO-domain protein catalyses thiazoline formation (204). In klebsazolicin biosynthesis, a single YcaO-domain protein catalyses formation of the 5-membered amidine ring as well as theazole heterocycles (205). In bottromycin, the macroamidine is formed in a one-step mechanism with direct nucleophilic attack of the N-terminal amino group (204). For klebsazolicin, Travin *et al* proposed two possible mechanisms for amidine ring formation in klebsazolicin: a one-step bottromycin-like mechanism (**Figure 9C**), or a two-step mechanism involving YcaO-dependent attack of the serine side chain adjacent to the amidine-forming residue, followed by attack of the N-terminal amino group and subsequent rearrangement to form the amidine ring (**Figure 9D**) (205).

1.2.3.5. Methylations

Methylation is a common post-translational modification found in multiple RiPP classes, where a methyl group is transferred to a carbon, oxygen or nitrogen atom in the metabolite. C-methylation occurs on many RiPPs, and is catalysed by radical S-adenosylmethionine (rSAM) enzymes (210). Two classes of rSAM enzyme have been implicated in RiPP biosynthesis: class B (B_{12} -dependent) and class C. Class B rSAM enzymes catalyse C-methylation in bottromycin (121), thiostrepton (211), polytheonamide (212) and siomycin (213), while class C enzymes catalyse C-methylation in nosiheptide (214), thiomuracin (215) and nocathiacin (214). The general rSAM-mediated mechanism involves production of a 5'-deoxyadenosyl (5'-dA) radical following cleavage of the SAM domain, which then abstracts a β -carbon hydrogen atom (216). Unusually, in thiostrepton A biosynthesis, the 5'-dA radical is not produced, and an external electron donor is not required for catalysis. Instead, a class B rSAM called TsrM catalyses the transfer of a methyl group from SAM to carbon-2 of tryptophan, during the first step of the quinaldic moiety ring expansion (211).

N-methylation is seen in RiPPs such as linaridins, LAPs and cyanobactins, and is catalysed by SAM-dependent methyltransferases. For example, cypemycin contains α -N-methylation at the N-terminus catalysed by the SAM-dependent methyltransferase CypM (217). Polytheonamides contain N-methylation on asparagine residues, catalysed by a SAM-dependent asparagine N-methyltransferase. This enzyme is highly promiscuous as it selectively methylates eight asparagine residues within the metabolite (177). In the biosynthesis of omphalotin A, which has a characteristic N-methylated backbone, a fusion protein OphMA autocatalytically and regioselectively methylates its own C-terminus (187,188). O-methylation is also catalysed by SAM-dependent methyltransferases following a similar mechanism, and is seen in bottromycins, LAPs and cyanobactins. SAM-dependent O-methyltransferases work by increasing the nucleophilicity of the target oxygen by lowering the pK_a , so that the electrophilic methyl group can then be transferred (218).

1.2.3.6. Lysine-to-tryptophan cross-linking

Lysine-to-tryptophan crosslinking is an unusual modification that is found in streptide and darobactin. The streptide biosynthetic gene cluster only encodes for one plausible modification enzyme, StrB, which is an rSAM enzyme. StrB was therefore hypothesised to catalyse the crosslinking between lysine and tryptophan, and this was confirmed through insertional mutagenesis of the *S. thermophilus* chromosome. StrB is thought to harbour only one auxiliary iron-sulphur [4Fe-4S] cluster, bound by its modified SPASM motif, in addition to the canonical SAM-activating cluster (174). A mechanistic model was proposed for streptide biosynthesis whereby reductive activation of SAM leads to formation of the 5'-dA radical, which abstracts a lysine β -hydrogen. The radical reacts with the indole side chain to create the crosslink and an indolyl radical. Deprotonation, rearomatisation and reduction of the auxiliary Fe-S cluster completes the synthesis of the crosslinked core peptide (174). The biosynthetic gene cluster for darobactin also encodes an rSAM enzyme, DarE, which is thought to be responsible for the lysine-to-tryptophan crosslinking in this metabolite. However, despite containing the SAM and SPASM domains, DarE has little overall homology with StrB. Additionally, DarE is thought to catalyse the unusual ether bond in darobactin, as there is no additional modification enzyme encoded in the *dar* operon that could carry out this reaction (95).

1.2.3.7. Lanthionine bonds

Lanthionine and methyllanthionine residues are installed onto serine and threonine residues via a two-step biosynthetic process. First, serine and threonine are dehydrated to Dha and Dhb respectively. Secondly, cysteine thiols are added to the unsaturated residues via a Michael-type addition to form (Me)Lan (219) (**Figure 10**). Different types

of enzyme have been found to catalyse dehydration and cyclisation of lanthipeptides, which form the basis for categorisation of lanthipeptide subclasses. LanB and LanC enzymes are involved in class I lanthipeptide biosynthesis, LanM for class II, LanKC for class III and LanL for class IV (220). LanB proteins are dehydratases that act via glutamation and elimination, and LanC is a cyclase. LanM is a fused synthetase that contains both an N-terminal dehydratase domain and a C-terminal cyclase domain. LanKC and LanL are both trifunctional enzymes containing an N-terminal lyase domain, a central kinase domain and a C-terminal cyclase domain (221,222). LanC and the cyclase domains of LanM and LanL contain a conserved zinc-binding motif (Cys-Cys-His/Cys), whereas the C-terminal domain of LanKC lacks these residues (223). In the recently described class V lanthipeptides, the corresponding gene clusters did not contain any of these typical lanthipeptide biosynthetic enzymes. Instead, dehydration and cyclisation are proposed to be catalysed by SprH3 and SprPT, enzymes that show homology to uncharacterised proteins in the thioviridamide biosynthetic pathway. As thioviridamide contains an AviCys moiety requiring serine dehydration, it was suggested that these enzymes might catalyse this modification in both cases (155).

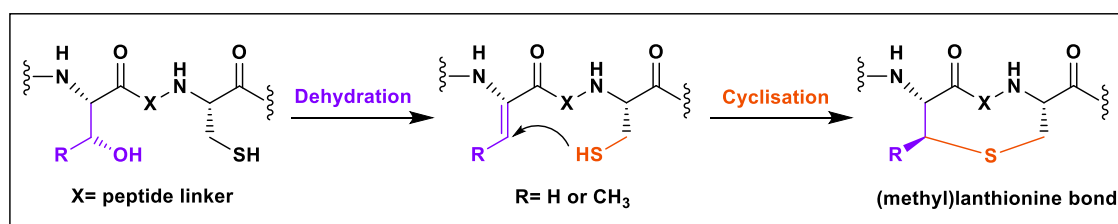


Figure 10: Mechanism of (methyl)lanthionine bond formation, which occurs during lanthipeptide biosynthesis (219).

1.2.3.8. Epimerisation

As RiPPs derive from standard proteinogenic amino acid building blocks, the residues are initially installed in L-configuration. However, many RiPP metabolites contain D-amino acids, which are introduced via epimerisation of α -carbons within core amino acids. As well as giving rise to further structural complexity, this alternative stereochemistry is thought to contribute to structural confirmation (224), bioactivity (225), and resistance to proteolysis (226). Incorporation of D-amino acids into natural products is usually achieved through either a deprotonation-protonation (227) or a radical mechanism (228) (**Figure 11**). In lanthipeptides, several other mechanisms have also been observed. For example, lactocin S contains D-alanine residues that are converted from genetically encoded L-serine residues (229). Based on the mechanism of (Me)Lan formation, the conversion of L-serine to D-alanine was suggested to occur via initial LanM-catalysed dehydration of L-serine to Dha followed by diastereoselective

hydrogenation of Dha by LanJ to yield D-alanine (230). This biosynthetic mechanism was also observed for lacticin 3147 (231). Epipeptides and proteusins also contain D-amino acids (216,232), which are introduced by an rSAM enzyme. Notably, polytheonamides contain 18 D-amino acids introduced by a single rSAM-like enzyme PoyD (233). rSAM-mediated catalysis follows abstraction of an α -hydrogen from carbon to form a stabilized amino acyl radical, followed by hydrogen donation from the opposite side of the residue to yield the epimer (116,216). In the recently discovered salinipeptides of the linaridin class, which contain nine D-amino acids, the genome of the producing organism does not contain the expected epimerases, suggesting that novel enzymology may be involved in producing these residues (164).

In some cases, azoline rings are capable of spontaneous epimerisation, which has been observed in bottromycin and cyanobactins (116). In bottromycin, an aspartate residue precedes the thiazoline ring, and it was proposed that only the D-form of aspartate is an appropriate substrate for the P450 decarboxylase enzyme. The P450-catalysed oxidation of the thiazoline locks the aspartate into D-form, as spontaneous epimerisation cannot happen preceding an azole ring (121). This is also observed in patellamides (234). In bottromycin biosynthesis, it was recently shown that BotH, an α/β -hydrolase fold enzyme, is responsible for the post-translational epimerisation of L-Asp to D-Asp, thus defining a new group of unusual peptide epimerases (235).

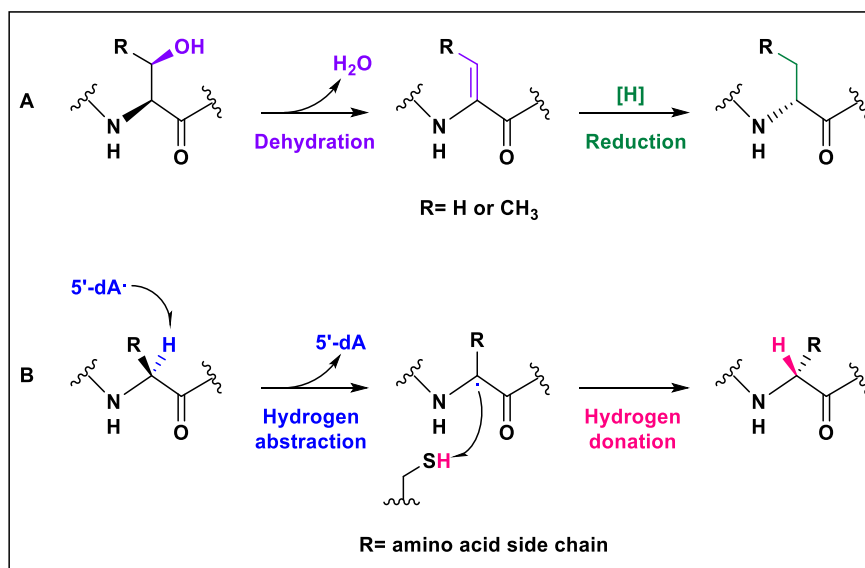


Figure 11: Mechanisms of epimerisation. A: epimerisation via dehydration and reduction of L-serine (R = H) or L-threonine (R =CH₃) to D-alanine (R =H) or D-2-aminobutyrate (R =CH₃) (227). B: Epimerisation via rSAM-mediated radical formation followed by hydrogen abstraction and donation (228).

1.2.3.9. Removal of leader peptide

A post-translational modification common to almost all RiPPs is cleavage of the leader and/or follower peptides to yield the mature RiPP metabolite. This is usually the final step of RiPP biosynthesis, as the leader and follower peptides are important for enzyme binding and core peptide stability. Leader peptide removal is usually catalysed by a dedicated protease from the biosynthetic pathway, but non-specific peptidases are also sometimes recruited. Overall, a range of different mechanisms of core peptide excision have been observed.

For biosynthesis of some LAPs such as microcin B17 and klebsazolicin, a 'molecular pencil sharpener' mechanism is adopted for leader peptide removal, exemplified by the TldD/E protease (236). This protease, a zinc or iron metalloprotein, is formed from two proteins that assemble into a spherical heterodimer. A narrow cleft in the centre then allows entry of unstructured peptides such as leader peptides, which are digested by an internal cleavage site (129,205). In the case of bottromycin, two separate proteases remove the C-terminal follower peptide and a single methionine residue at the N-terminus of the precursor peptide. The follower peptide is cleaved by an aminohydrolase, and the methionine is removed by a methionine aminopeptidase (121,237). For thiopeptides that contain pyridine rings, leader peptides are removed by elimination rather than hydrolysis. The [4+2] Diels-Alderase catalyses elimination of the leader peptide as a carboxamide, resulting in the subsequent dehydrogenation required to form the pyridine ring (238).

In lanthipeptide biosynthesis, a LanP protease is adopted to cleave leader peptides, which is sometimes coupled with export of the final metabolite (230). For class I and class II lanthipeptides, a subtilisin-like serine peptidase is often recruited for leader peptide cleavage (230). In other cases, a bifunctional LanT_P enzyme from the ABC transporter maturation and secretion family catalyses proteolysis followed by export (239,240). Cinnamycin is an unusual example of a class II lanthipeptide whose leader peptide is removed by a protease from the general secretory (Sec) system (241). In fact, many lanthipeptide gene clusters from streptomycetes appear to lack lanthipeptide-specific proteases (242). Proteases for class III and IV lanthipeptides are less well characterised. In some cases, endogenous proteases catalyse leader peptide digestion (243), whereas in other cases a prolyl oligopeptidase selectively cleaves leader peptides from the mature metabolite (244). Recently, it was shown that a bifunctional zinc-dependent protease is responsible for processing of class III lanthipeptides, and homologues of this protein were also found in class IV lanthipeptide pathways (245). For

class V lanthipeptides, an M16-domain peptidase is thought to excise the mature RiPP (155).

1.2.3.10. Biosynthesis of cyclic plant RiPPs

During cyclotide biosynthesis, disulphide bonds are formed by protein disulphide isomerases in the endoplasmic reticulum of the plant cell (246). The modified core peptide is then cleaved at the N- and C-terminus and cyclised by an asparagine-specific endopeptidase in the plant vacuole (247). Orbitides are produced in a similar way, with endoproteolytic cleavage of the N-terminus of the core peptide followed by C-terminal proteolysis and cyclisation catalysed by serine proteases (248,249). Lyciumin precursor peptides contain a C-terminal BURP domain (193), which is usually associated with abiotic stress response in plants (250). The precursors also contain multiple core sequences, producing several lyciumin analogues. The core peptides are cyclised between each tryptophan and glycine via a radical-oxidative cyclisation mechanism, and then cleaved at the N-terminus by an endopeptidase. Finally, the peptides are protected at the N-terminus by pyroglutamate formation, which can be catalysed by a glutamine cyclotransferase (193).

1.3. Thesis Aims

1.3.1. Importance of RiPP Discovery

RiPPs are a structurally diverse class of natural product and are of particular interest for their potential as chemotherapeutic agents, with many antibiotic RiPPs having recently been reported (95,129,154,165,171,172,251). Moving forward, it is important to identify metabolites with novel structural scaffolds, as antibiotic resistance mechanisms will evolve more easily to metabolites that are structurally similar to those currently in clinical use. Additionally, highly modified metabolites are much less susceptible to proteolysis *in vivo*, improving their pharmacokinetic potential. Therefore, the chemical complexity of RiPPs makes this class an exciting group of metabolites to explore for future drug candidates. The discovery of novel RiPPs also enables further study of their biosynthesis, helping to elucidate how and why these fascinating natural products are produced so widely in nature.

1.3.2. Gaps in Current Understanding

Previous analyses of bacterial genomic data have demonstrated that there is a considerable amount of untapped biosynthetic diversity within the RiPP landscape (124,251–253). It is also widely appreciated that numerous natural product biosynthetic gene clusters are transcriptionally silent or ‘cryptic’ under laboratory conditions, and many of these remain uncharacterised. Additionally, although widely studied, there are several aspects of RiPP biosynthesis that are still unknown. Particularly, YcaO-domain proteins are widely present in bacterial genomes, but the exact biosynthetic role of many of these enzymes remains unknown (124,197,198).

1.3.3. Aims and Objectives of Thesis

The overall aim of this thesis is to investigate the unexplored diversity of RiPPs by genome mining for novel RiPP biosynthetic gene clusters and characterising their products. This research project is based on the following objectives:

- (i) Identify novel RiPP biosynthetic gene clusters associated with YcaO-domain proteins
- (ii) Characterise a novel RiPP following genetic manipulation of a model pathway

1.3.4. Thesis Outline

Chapter 2 discusses RiPP genome mining and describes the discovery of novel gene clusters using a newly developed mining tool. Chapters 3 and 4 describe the genetic, metabolomic, structural and biological analyses carried out to characterise a novel RiPP. Chapter 5 will discuss the wider implications of these results on future RiPP discovery.

Chapter 2: Genome Mining for Novel RiPP Gene Clusters

2.1. Introduction

2.1.1. Genome Mining and the Genomic Era

As discussed in Chapter 1, microorganisms and plants produce a plethora of specialised metabolites which have important uses in medicine, agriculture and the food industry (3,5). In order to discover more of these useful metabolites, one approach that has been increasingly utilised in natural products discovery is genome mining. Genome mining is a computational method for the automatic detection and annotation of biosynthetic gene clusters from genomic data (254).

The first bacterial genome was sequenced in 1995 from *Haemophilus influenzae* (255), a Gram-negative coccobacillus and human pathogen that causes diseases such as pneumonia, meningitis and otitis media (256). This historic sequencing project was led by Craig Venter at The Institute for Genomics Research in Rockville, USA. Seven years later in 2002, the first *Streptomyces* genome was published from *Streptomyces coelicolor* A3(2) (257). This was a huge sequencing effort led by Sir David Hopwood from the John Innes Centre in collaboration with the Sanger Institute in Cambridge. This genome sequence gave the first insights into the fascinating genetics and biosynthetic potential of streptomycetes, and *S. coelicolor* A3(2) remains an important model organism for *Streptomyces* research today. This genome was shortly followed by that of *Streptomyces avermitilis* in 2003 (258), a project led by Satoshi Ōmura in Japan. *S. avermitilis* is an important industrial microorganism that produces the antimalarial avermectin metabolites (258). The analysis of these early streptomycete genomes indicated that actinomycetes contain many more gene clusters for specialised metabolites than previously thought (259), with both *S. coelicolor* and *S. avermitilis* genomes encoding ~10 times the number of natural products than had been identified from culture fermentations. An analysis by Baltz suggested that actinomycetes encode 25-50 biosynthetic gene clusters for secondary metabolites, a number that increases with the size of the organisms genome (260).

In the last 25 years, sequencing technologies have become increasingly more advanced, accessible and cheaper. As a result, the number of prokaryotic genomes that are publicly

available exceeds 250,000 (**Figure 12**), of which over 23,000 are from Actinobacteria (NCBI, May 2020).

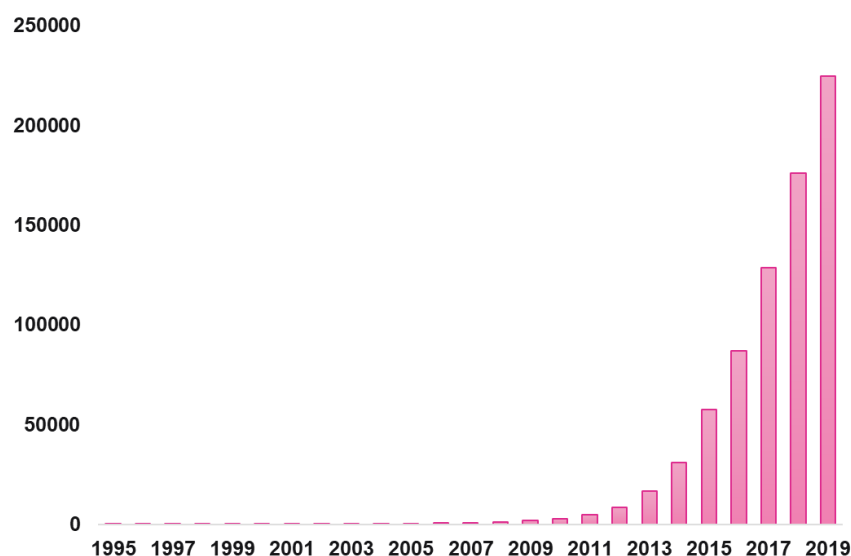


Figure 12: Cumulative number of prokaryotic genomes deposited in the NCBI database. Data collected between 1995 and 2019. Source: NCBI genome reports, downloaded February 2020.

This wealth of sequence information paved the way for numerous bioinformatic platforms to aid biological research such as genome visualisation software (261–263), alignment tools (264,265), homology search platforms (266) and prediction tools for open reading frames and protein functions (267,268). Databases of known natural products have also been put together, such as NP atlas. This contains key structural data and information about the origin of many known RiPP metabolites. Several genome mining tools and databases of biosynthetic gene clusters have also been developed. Genome mining tools comprise algorithms that are based on knowledge of natural product biosynthetic machinery, which are used to survey genomic data and identify novel biosynthetic pathways. As well as identifying new metabolites from microorganisms that are known to be talented producers of natural products, genome mining is also a valuable tool to help understand the biosynthetic potential of underexplored genera. Overall, there is great potential for genome mining strategies to revitalise the antibiotic pipeline, at a time when discovery rates are dwindling and antimicrobial resistance is increasing (24).

2.1.2. Genome Mining for RiPP Biosynthetic Gene Clusters

As outlined in Chapter 1, ribosomally synthesised and post-translationally modified peptides (RiPPs) are a structurally complex class of natural product produced from diverse biosynthetic machineries (116,117). Due to the biosynthetic logic of RiPPs, genome mining for novel RiPP gene clusters presents several challenges. Unlike other classes of natural product such as polyketides and nonribosomal peptides that are produced by multi-modular complexes, there are very few conserved biosynthetic features across the RiPP class. The RiPP tailoring enzymes responsible for modifications vary between different RiPP subfamilies. Furthermore, the precursor peptides are very small, typically fewer than 100 amino acids, and are rarely annotated in genomes. As a result, many genome mining tools that were initially developed for natural product discovery were less successful at identifying RiPP gene clusters compared to other classes. More recently, several further genome mining tools have been developed that are optimised for RiPP discovery. These comprise algorithms built from the ever-increasing knowledge of RiPP tailoring enzymes and precursor peptide sequences. The use of more bespoke genome mining tools represents a powerful strategy to discover previously untapped biosynthetic diversity within this natural product class.

The RiPP genome mining tools that have been developed can take a variety of different inputs as a starting point for analysis, including whole genomes, precursor peptide sequences and individual tailoring enzymes. These are then analysed to identify and annotate associated RiPP gene clusters. A number of tools also provide additional outputs such as prediction of precursor peptide sequences, leader peptide cleavage sites, post-translational modifications and molecular structures. As well as genomic analysis, some tools also analyse and integrate mass spectrometry data. RiPP genome mining tools that are currently available are summarised in the following sections, highlighting their various features. These are described in order of when they were released.

2.1.2.1. Analysis of whole genomes to identify RiPP biosynthetic gene clusters

2.1.2.1.i. BAGEL

BAGEL (BActeriocin GEnome mining tool, <https://bagel4.molgenrug.nl/>) was the first genome mining tool built with specific rules for RiPP and bacteriocin identification, released in 2006 (269). BAGEL works by first identifying areas of interest in the genome based on information about the genetic context of accessory genes. Then, small open reading frames are searched for within these regions, and blasted against knowledge-based peptide and motif databases (269). Since its first release, software updates have

provided further optimisation for RiPP identification, such as updated peptide and motif databases (270) and extended Hidden Markov Models (HMMs) for RiPP classes such as cyanobactins, sactipeptides and linaridins (271). The latest version, BAGEL4, was updated with improved RiPP protein domain information (272).

2.1.2.1.ii. *antiSMASH*

One of the most widely used genome mining tools is antiSMASH (antibiotics and Secondary Metabolite Analysis Shell, <https://antismash.secondarymetabolites.org>), which has processed over 750,000 jobs at the time of writing. antiSMASH has the capacity to identify 52 types of natural product biosynthetic gene cluster, including those for RiPPs. antiSMASH works by comparing encoded gene products with a library of profile HMMs (pHMMs), which describe a range of biosynthetic genes. Gene clusters are then identified by assigning key enzymes to “specialised metabolism clusters of orthologous groups” (273). Further analyses are then carried out to annotate accessory genes and predict cluster boundaries, substrate specificity and molecular structures (273,274). antiSMASH was first released in 2011 (273) and has since been updated several times (274–277), incorporating improved detection rules for RiPPs including thiopeptides, sactipeptides, lanthipeptides, lasso peptides, LAPs and radical SAM-associated RiPPs. Many genome mining tools are limited to bacterial genomes, but iterations of the antiSMASH software have also been developed for fungal (fungiSMASH (276)) and plant (plantiSMASH (278)) genome mining.

2.1.2.1.iii. *RiPP-PRISM*

RiPP-PRISM (<http://grid.adapsyn.com/prism/#!/prism>) is a genome mining tool that identifies gene clusters of 21 RiPP families (252). It was first reported in 2013, integrating with the previously developed PRISM (PRediction Informatics for Secondary Metabolomes) software (279,280), which identifies non-ribosomal peptide and polyketide gene clusters. RiPP-PRISM comprises libraries of motifs, HMMs and putative tailoring modifications specific to RiPPs, which are used to predict precursor peptide cleavage sites and molecular structures (252).

2.1.2.1.iv. *RiPPMiner*

RiPPMiner (http://202.54.226.242/~priyesh/rippminer2/new_predictions/index.php), released in 2016, is a tool that identifies 13 different families of RiPP gene cluster and helps to predict molecular structures for metabolites such as lanthipeptides, lasso peptides, cyanobactins and thiopeptides (281). RiPPMiner comprises predictive power from support vector machine and random-forest classifiers trained on over 500 experimentally characterised RiPPs. These are used to discriminate genuine precursor

peptides from other small peptides, and classify these into a particular subfamily. RiPPMiner encompasses two different platforms for genome and peptide analysis. RiPPMiner-peptide analyses precursor sequences providing predictions about class, structure, crosslinks and cleavage sites. RiPPMiner-genome identifies gene clusters from a genomic sequence and predicts associated chemical structures.

2.1.2.1.v. DecRiPPter

DecRiPPter (Data-driven Exploratory Class-independent RiPP TrackER) (155) is a recently developed tool that combines machine learning and pan-genomic analysis to identify novel RiPP biosynthetic gene clusters. A Support Vector Machine identifies putative RiPP precursor peptides from the accessory genome of a genus to identify those that are encoded within operon-like pathways. Gene clusters are then prioritised based on the presence of novel enzymology and patterns of biosynthetic gene conservation across species (155).

2.1.2.2. Mass spectrometry-guided mining tools for RiPPs

Mass spectrometry (MS) has become an increasingly powerful method in natural product research and discovery (282). As a result, MS-based analyses have been integrated into several recently developed genome mining tools to improve detection of novel metabolites. In the context of peptidic natural products such as RiPPs, analysing metabolomic data can provide useful clues about amino acids present in a metabolite based on molecular fragmentation patterns from MS/MS data. Additionally, post-translational modifications present in a metabolite can correspond to characteristic mass losses, providing further structural clues.

A notable tool that aids MS-based natural product discovery is Global Natural Products Social (GNPS) molecular networking (<https://gnps.ucsd.edu/>) (100). This tool uses tandem MS (MS/MS) to identify families of related metabolites in spectra the user uploads, and compares this to a large database of MS/MS spectra. Networks of related metabolites are created which can be visualised to help identify connections within metabolomic datasets. GNPS has also opened up the potential to utilise a vast amount of publicly available metabolomic datasets for natural product discovery, which are uploaded by the community. Since its release, over 1,500 datasets have been contributed to GNPS (as of April 2020), corresponding to 35 terabytes of data. A search for common RiPP compound names in the 'molecular explorer' function of GNPS did not reveal any spectral matches to the library, suggesting that these metabolites are not currently commonly annotated in mass spectral networks. However, searching the

GNPS-MassIVE dataset repository reveals some metabolic datasets for RiPPs such as bottromycin and thiostrepton.

2.1.2.2.i. Natural product peptidogenomics

One of the first examples of MS-guided genome mining was reported by Kersten *et al.* in 2011, who developed Natural Product Peptidogenomics (NPP) (283). This tool aimed to connect chemotypes of peptide natural products with their biosynthetic genes, combining advances in mass spectrometry, genomics and knowledge of natural product biosynthesis. NPP works by carrying out an initial MALDI-TOF MS analysis, which searches for uncharacterised small peptide masses between 1,500-5,000 Da. Putative peptides are picked out based on fragmentation patterns, which are used to define sequence “search tags”. These are then compared to a six-frame translation of the genome to identify putative precursor peptides. Further analytical steps use biosynthetic knowledge to ensure that fragmentation-based connections to genome-derived structures makes sense biosynthetically.

2.1.2.2.ii. RiPPquest and MetaMiner

RiPPquest, released in 2014, is another tool that combines metabolomic and genomic data to discover RiPPs, with particular optimisation for lanthipeptides (284). One disadvantage of the NPP approach is that macrocyclic RiPPs might be overlooked by the sequence search tag method, and so RiPPquest was built to overcome this. RiPPquest works by first predicting lanthipeptide biosynthetic gene clusters and putative precursor peptides from a genomic input. MS/MS spectra for all potential lanthipeptide structures are calculated based on conceivable post-translational modifications on putative core peptides. Then, peptide-spectrum matches are scored in order to connect metabolomic and genomic data. Finally, a molecular network is generated from the fragmentation data, in order to identify homologues of characterised lanthipeptides and families of related peptides.

Although a useful tool, RiPPquest was limited to the discovery of lanthipeptides from small datasets with predefined sets of post-translational modifications. An updated tool, MetaMiner, was released in 2019 (285), which expanded RiPP discovery to search for lanthipeptides, LAPs, lasso peptides, linaridins, glycocins, cyanobactins, proteusins, phenol-soluble modulins and auto-inducing peptides. MetaMiner is integrated into GNPS (<http://gnps.ucsd.edu/ProteoSAFe/static/gnps-theoretical.jsp>) as well as the Natural Product Discovery tools package on Github (<https://github.com/ablab/npdtools>). MetaMiner works by first analysing the paired genome/metagenome assemblies and fragmentation data from a given input. From this, putative biosynthetic gene clusters and

associated precursor peptides are identified using antiSMASH (273) and Bacteriocin Operon and gene block Associator (286). Target and decoy putative RiPP structure databases are then constructed, which are used to search for likely precursor peptides. Tandem mass spectra are then compared against these databases, and mass spectral networking is employed to expand the set of described RiPPs.

2.1.2.2.iii. *CycloNovo*

Recently in 2020, CycloNovo was reported, a tool for the discovery of cyclic peptides including cyclic RiPPs from large datasets (287). The software is integrated into GNPS (<https://gnps.ucsd.edu/ProteoSAFe/index.jsp?params=%7B%22workflow%22:%22CYCLONOVO%22%7D>) as well as the Natural Product Discovery tools package on Github (<https://github.com/bbehsaz/cyclonovo>). Fragmentation of cyclic peptides is more unpredictable than for linear peptides, as it can occur at any amide bond producing a complex series of ions that might not necessarily match the primary amino acid sequence. CycloNovo overcomes this problem by using de Bruijn graph representations of spectra. This involves calculation of putative k-mers (strings of k consecutive amino acids) for putative cyclopeptides, which are then scored against input spectra (287). Cyclospectra that are identified by CycloNovo can be further analysed with GNPS-hosted tools such as Dereplicator/Varquest (288,289) to construct and annotate molecular networks.

2.1.2.2.iv. *DeepRiPP*

Another tool reported in 2020 is DeepRiPP (<http://deepripp.magarveylab.ca/>), which combines a range of genomic and metabolomic information to automate discovery of biosynthetic gene clusters and associated RiPP structures (290). DeepRiPP comprises three platforms that can be used individually or as a combined full workflow. The first platform is NLPPrecursor, a deep neural network (DNN)-based tool which identifies precursor peptides and predicts their RiPP class and cleavage sites. The second platform comprises the BARLEY (Basic Alignment of Ribosomal Encoded Products Locally) algorithm, which compares biosynthetic loci to known RiPP gene clusters. This deduces post-translational modifications within the biosynthetic gene cluster and compares the putative RiPP product with a database of known RiPPs. A similarity score is provided between the identified gene cluster and RiPP database, which thereby aims to prioritise RiPP novelty. The third platform, CLAMS (Computational Library for Analysis of Mass Spectra), employs an algorithm that compares mass spectrometry data with candidate RiPP gene clusters. This works by taking the exact mass of a predicted RiPP and searching for supporting fragmentation patterns within the MS/MS data.

2.1.2.3. Bespoke approaches to RiPP genome mining

2.1.2.3.i. RODEO

RODEO (Rapid ORF Description and Evaluation Online, <https://ripp.rodeo/index.html>) is a tool released in 2017 that analyses RiPP gene clusters and predicts precursor peptides and molecular structures (253). Unlike the previously developed tools that use whole genomes or MS data as input, RODEO centres its analysis on a protein of interest, capturing the surrounding genomic environment. Within this genomic region, nearby biosynthetic genes are identified, piecing together a gene cluster. The RODEO algorithm combines pHMM and motif-based analysis, heuristic scoring and machine learning to identify precursor peptides and predict cleavage sites. RODEO was initially built for lasso peptide identification, but the software has also been subsequently optimised for thiopeptide and sactipeptide recognition (251,253,291).

2.1.2.3.ii. NeuRiPP

NeuRiPP (<https://github.com/emzodls/neuripp>), reported in 2019, is a tool that identifies precursor peptides and distinguishes genuine RiPP precursors from false positives (292). This approach might be useful if a given tool identifies multiple putative precursor peptide sequences. NeuRiPP comprises a DNN trained on positive and negative precursor peptide datasets of over 9,000 sequences. The positive dataset included experimentally validated sequences from tools including PRISM (279), ThioFinder (293), RODEO (253), RiPPER (124) and antiSMASH (277), while the negative dataset comprised sequences known not to be genuine RiPP precursors. Once trained, the DNN was used to categorise short sequences on their likelihood of being genuine RiPP precursor. NeuRiPP has been successful at identifying precursor peptides enriched with HMMs for known RiPPs, as well as precursors for RiPP classes it was not trained on, with over 99% accuracy.

2.1.2.3.iii. RRE-Finder

Another recently developed tool, RRE-finder (294), was built for the identification of RiPP recognition elements (RREs), which can then be used to help discover novel RiPP classes. RRE-Finder has two modes of operation: “precision” and “exploratory”. Precision mode utilises 35 custom pHMMs designed to detect RRE domains in a class-dependent manner. These pHMMs are built from known RRE-containing proteins that have been verified to bind their cognate precursor peptide. Precision mode can be used to predict the presence of an RRE domain as well as the RiPP class that the cognate precursor peptide belongs to. Exploratory mode uses a shortened version of the HHpred workflow (295) with a custom database of detected RREs. This mode can be used to

identify a wider range of putative RRE-containing proteins to assist in the discovery of novel RRE-dependent RiPP classes (294).

2.1.2.4. Application of recently developed genome mining tools to discover novel RiPPs

The genomic and MS-based tools described above have been widely applied to microbial genomes, which has led to the discovery of several new natural products with important bioactivities and novel structural features. Furthermore, the application of these tools has demonstrated the vast number of RiPP biosynthetic gene clusters present in genomes that were previously unknown, identifying new RiPP families and highlighting that RiPPs occupy a much larger genetic space than previously appreciated, and are produced in a range of different environments.

2.1.2.4.i. Expanding RiPP chemical space

In 2013, RiPP-PRISM was applied to over 65,000 prokaryotic genomes leading to identification of over 30,000 previously uncharacterised RiPP gene clusters. This analysis suggested that, at the time, at least 82% of genetically encoded RiPPs were uncharacterised (252). In 2017, RODEO was used to survey the genomic space occupied by lasso peptides, leading to the identification of 1,400 lasso peptide gene clusters (253). In 2018, RODEO analysis expanded the thiopeptide class by a factor of four (251) and further updates to the software led to discovery of a new RiPP family related to sactipeptides, called the ranthipeptides (radical non- α thioether peptides) (291). In 2020, DeepRiPP was used to investigate the presence of lanthipeptide gene clusters in over 65,000 bacterial genomes, leading to the identification of over 19,000 novel RiPPs (290). The recently developed decRiPPter tool analysed 1,295 *Streptomyces* genomes, leading to the identification of 42 putative new RiPP families that could not be found with existing genome mining programmes (155).

In terms of MS-based discovery, MetaMiner was applied to mass spectrometry databases to identify seven unknown RiPPs in some unusual sources including the sponge microbiome, the International Space Station and the human microbiome, demonstrating that these tools can be applied to diverse datasets (285). CycloNovo identified over 400 previously unreported cyclic peptides from GNPS datasets and was also used to analyse a human stool dataset leading to the discovery of several bioactive cyclic peptides that had remained stable throughout the gastrointestinal system. This was the first indication that such metabolites can survive these conditions (287).

2.1.2.4.ii. Discovery of structurally novel RiPPs

As well as assessing the genetic space occupied by RiPPs, many of these genome mining tools have also led to the isolation of novel metabolites. antiSMASH analysis guided the discovery of a novel class IV lanthipeptide streptocollin (296). RiPP-PRISM analysis led to the characterisation of aurantizolicin from *Streptomyces aurantiacus*, a cyclic azoline-containing metabolite closely related to the YM-216391 family (252) (**Figure 13**). RODEO analysis led to the discovery of novel lasso peptide from *Nocardiosis alba* which forms a novel 'handcuff' topology due to co-location of two cysteine residues on the lasso tail. A further lasso peptide was isolated from *Streptomyces albulus* NRRL B-3066, called citrulassin A, which harbours a unique post-translational modification where a genetically encoded arginine residue is modified to citrulline (253). RODEO analysis also identified a novel thiopeptide called saalfelduracin from *Amycolatopsis saalfeldensis* NRRL B-24474, which contains a structurally unique piperidine and a rare thioamide moiety. Saalfelduracin also exhibited strong antibacterial activity against VRE and MRSA (251). A novel sactipeptide called huazacin was also isolated following RODEO analysis, which displayed growth-suppressive activity against *Listeria monocytogenes* (291). RODEO analysis has also identified novel lanthipeptides, such as LP2006 (291) (**Figure 13**). RiPPquest analysis led to characterisation of a new class II lanthipeptide called informatipeptin from *Streptomyces viridochromogenes* (284), and DeepRiPP analysis led to identification of deepstreptin, a lasso peptide, and two lanthipeptides called deepflavo (**Figure 13**) and deepginsen (290). The application of decRiPPter to *Streptomyces* genomes led to the discovery of the novel class V lanthipeptide subfamily (155).

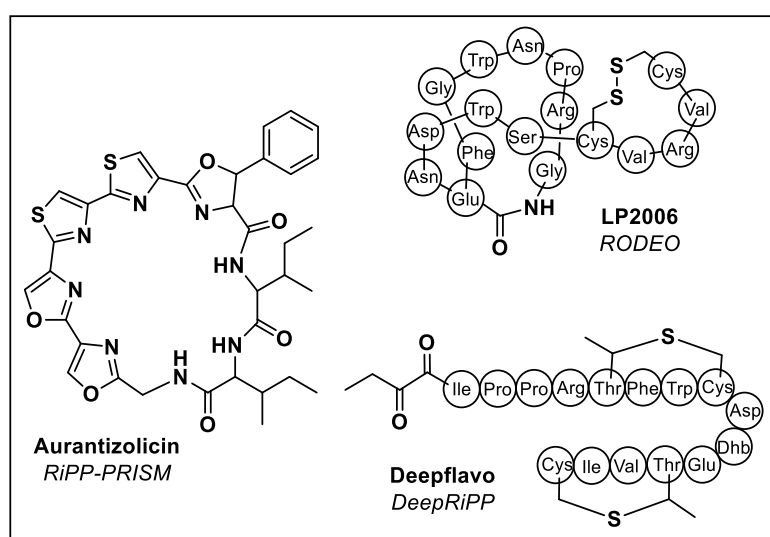


Figure 13: Examples of RiPP natural products discovered by genome mining. LP2006 is a novel lanthipeptide discovered from a RODEO analysis (291), aurantizolicin is an azoline-containing metabolite discovered from a RiPP-PRISM analysis (252) and deepflavo is a lanthipeptide discovered from a DeepRiPP analysis (290).

2.1.3. RiPPER: a Novel Gene-led Genome Mining Tool for RiPPs

Whilst some of the metabolites described above harbour structural novelty, many of the RiPPs discovered using currently available tools are limited to known RiPP classes. This is partly due to the fact that they are trained on sequence information and biosynthetic logic of known RiPPs. Therefore, the opportunity to identify truly novel RiPP metabolites with unique structural and biosynthetic features might be missed. RODEO is one tool that has been particularly successful at identifying structural and biological novelty, whose algorithm analyses genomic regions centred on a target protein. This suggests that a bespoke gene-led approach might be a promising tool for novel RiPP discovery. With this in mind, a new genome mining tool was developed by Andrew Truman and Govind Chandra at the John Innes Centre, which aimed to provide unbiased RiPP precursor peptide detection. This tool was called RiPPER (RiPP Precursor Peptide Enhanced Recognition) (124).

2.1.3.1. *RiPPER workflow*

RiPPER takes a known or putative RiPP tailoring enzyme as an input, and uses this to capture surrounding biosynthetic genes and annotate short peptides that could be novel precursor peptides (124). To achieve this, RiPPER uses the RODEO2 (251) script to capture genomic regions surrounding the 'bait' tailoring gene. A modified version of Prodigal (297) called Prodigal-short is then employed to re-annotate the captured biosynthetic gene cluster for short protein-coding open reading frames that could encode RiPP precursor peptides. The peptides with the highest Prodigal-short scores are retrieved and assessed for several characteristics, including conserved domains such as Pfam domains and RiPP-specific HMMs from NCBI. The annotated genomic regions captured by RiPPER can be visualised with Artemis (**Figure 14**) (261), where the 'bait' gene is highlighted in green, and putative precursor peptides are highlighted in red. A colour scale of light to dark red is used to indicate the highest scoring precursor peptides, and a list of their scores is also shown.

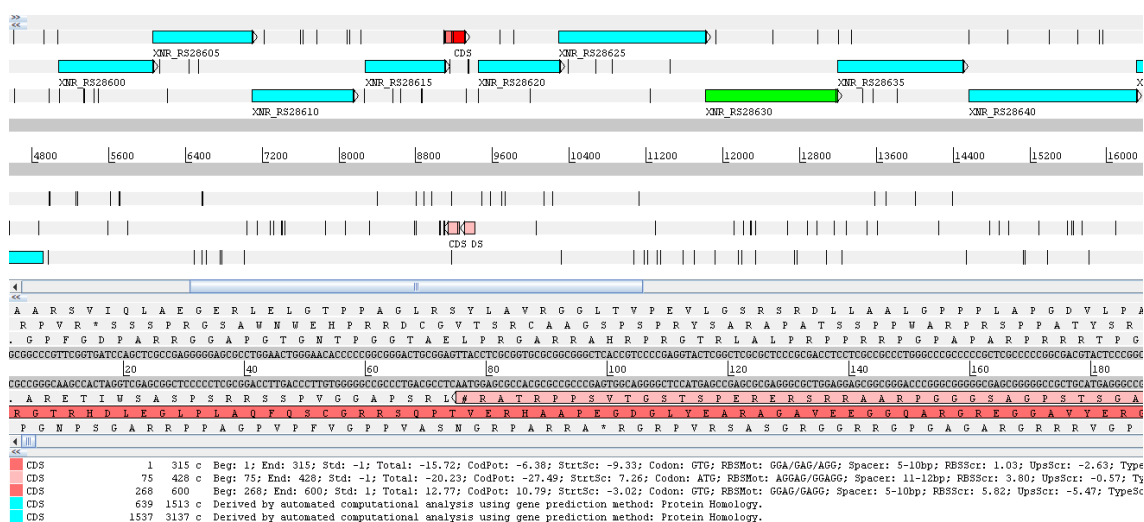


Figure 14: RiPPER output of an annotated microbial genome viewed with Artemis. The bait protein that the analysis is centred on is highlighted in green and putative precursor peptides are highlighted in varying intensities of red based on their precursor peptide “score”, which is calculated based on features such as GC content and presence of ribosome binding sites. Existing knowledge of RiPP biosynthesis can also be used to help validate the gene clustering observed in the RiPPER output.

For each tailoring enzyme submitted, the three highest scoring nearby precursor peptides are retrieved by RiPPER. The identified precursor peptides can then be analysed via molecular networking using EGN (Evolutionary Gene and genome Network) (298), which helps to identify families of related precursor peptides. RiPPER is therefore useful for analysis of multiple related biosynthetic gene clusters and precursor peptides, and was shown to identify families of lasso peptides, thiopeptides and microviridins without any prior knowledge of their precursor peptide sequence motifs. Due to the use of user-defined protein accessions as a starting point for analysis, RiPPER is a flexible tool that can be applied to various RiPP classes and can be used to identify precursor peptides that have no homology to known families of RiPP, as well as those with known RiPP precursor domains. RiPPER also provides an accurate re-annotation of genomic loci for small genes missed by automated genome annotations.

2.1.3.2. RiPPER-led discovery of the thiovarsolins

RiPPER was successfully used by Santos-Aberturas *et al* to study the unexplored diversity of thioamidated RiPPs, which was achieved by using an input of TfuA-like proteins from Actinobacteria (124). As discussed in Chapter 1, TfuA proteins are partners to YcaO-domain proteins, which together catalyse thioamidation in the peptide backbone of RiPPs such as thiopeptides (207,251) and thioamitides (150,151,208). This RiPPER analysis led to the retrieval of 743 precursor peptides nearby bait TfuA proteins, which were subsequently grouped into 74 distinct precursor networks (124) using EGN (298). A model biosynthetic gene cluster from one of these networks was characterised through

cloning and heterologous expression of the pathway from *Streptomyces varsoviensis*. Four metabolites produced by the gene cluster were identified by metabolomic analysis, which were isolated and characterised as the thiovarsolins A-D (**Figure 15**). These metabolites feature three interesting post-translational modifications: thioamidation, N-acetylation and an unusual dehydrogenation. These metabolites thus describe a new structural class of thioamidated RiPP (124).

The RiPPER-led discovery of the thiovarsolins, as well as several other networks of unknown RiPP precursor peptides, shows that this genome mining approach is successful at identifying previously uncharacterised RiPP gene clusters using a bait RiPP tailoring enzyme as an input.

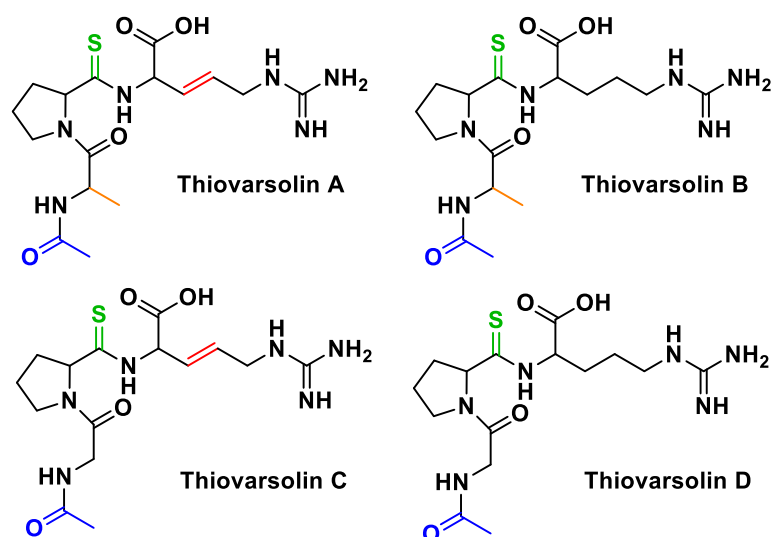


Figure 15: Chemical structures of thiovarsolins A-D, novel RiPPs identified through genome mining using RiPPER. Post-translational modifications are highlighted on each structure: thioamide bonds (green), dehydrogenation (red), N-acetylation (blue). Thiovarsolins A and B also contain an additional methyl group (highlighted in orange) compared with thiovarsolins C and D, due to the presence of alanine instead of glycine in the core peptide.

2.2. Chapter Aims

RiPPs are a largely underexplored class of natural product, partly due to the challenges encountered when genome mining for RiPP pathways. The aim of this chapter is to examine the presence of YcaO-domain proteins in microbial genomes and use these to guide the discovery of novel RiPP biosynthetic gene clusters. To achieve this, the specific objectives of this chapter were:

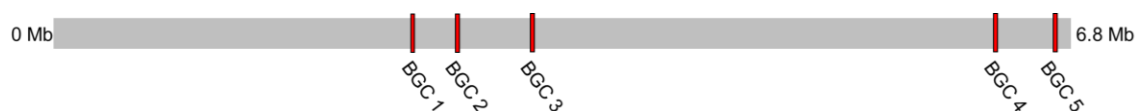
- (i) Carry out a comparative analysis of genome mining tool outputs
- (ii) Use RiPPER to retrieve precursor peptides associated with YcaO-domain proteins
- (iii) Carry out bioinformatic analyses of identified precursor peptides and associated biosynthetic gene clusters

2.3. Results and Discussion

2.3.1. Comparative Analysis of Genome Mining Outputs from the Genome of *Streptomyces albus* J1074

In order to compare the outputs of different genome mining tools, I used the genome sequence of *Streptomyces albus* J1074 as input for tools that analyse whole genomes (antiSMASH5, BAGEL4, DeepRiPP, PRISM4 and RiPPMiner). I chose this organism as it is a widely used model streptomycete and its biosynthetic capacity is well-studied (299,300), therefore it would be interesting to see the differences in gene cluster identification by the different tools.

Considering the outputs of all five tools (**Appendix Figure 78****Figure 82**), five distinct RiPP biosynthetic gene clusters were identified in total (**Figure 16**). Surprisingly, only two of these five clusters were identified by all five tools: BGC2 and BGC5, which are both lanthipeptide gene clusters. This suggests that these genome mining tools harbour good predictive power for lanthipeptide RiPPs, although interestingly the tools varied in their description of the precise lanthipeptide class. For BGC2, all five tools provided a predicted precursor peptide sequence, but these peptides also varied across the outputs. antiSMASH, DeepRiPP and PRISM identified the same precursor peptide sequence, but only antiSMASH and PRISM highlighted the predicted core peptide. antiSMASH also highlighted residues that are dehydrated to Dha or Dhb. BAGEL identified a similar precursor peptide but with an earlier alternative start codon, but also highlighted the same core peptide as PRISM. RiPPMiner provided a much longer precursor peptide with a short core peptide, which is likely to be incorrect. For BGC5, only DeepRiPP, PRISM and RiPPMiner provided precursor peptide sequences for the lanthipeptide, but these sequences were all different. This highlights the variability of precursor peptide prediction across different tools, even when a similar biosynthetic gene cluster is identified. BGC1 is a LAP/thiopeptide cluster identified by all tools except BAGEL. The same precursor peptide sequence was provided by DeepRiPP, PRISM and RiPPMiner, but DeepRiPP predicted a shorter core peptide compared to the other two tools. BGC3 is for a hypothetical linaridin, which was only identified by RiPPMiner. Similarly, BAGEL was the only tool to identify BGC4, a linocin M18-like bacteriocin cluster. This demonstrates that utilising different tools with the same genome can be a useful way to identify otherwise unidentifiable gene clusters. As well as the RiPP gene clusters, antiSMASH identified 19 other natural product biosynthetic gene clusters and PRISM identified 12 other natural product gene clusters from the genome of *S. albus* J1074.



BGC	Tool	Description	Predicted precursor peptide
1	antiSMASH	LAP/Thiopeptide	
	BAGEL	Not identified	
	DeepRiPP	Thiopeptide	MDPWDRREYAMTPKTELATLADEILELESETFEISDYSDAAEVVL AGST SCSSTSCSSTTSTSCSA
	PRISM	Thiopeptide	MDPWDRREYAMTPKTELATLADEILELESETFEISDYSDAAE VVL AGSTSCSSTSCSSTTSTSCSA
	RiPPMiner	LAP/Thiopeptide	MDPWDRREYAMTPKTELATLADEILELESETFEISDYSDAAE VVL AGSTSCSSTSCSSTTSTSCSA
2	antiSMASH	Lanthipeptide	MALLDLQAMDTPQEEAVGDLAT GDhaQIDhaLLICEYDhaDhaL DhaVDhblCDhbP
	BAGEL	Lanthipeptide class II	VAVVRPLDTKEYVMALLDLQAMDTPQEEAVGDLA TGSQISLLIC EYSSLSVTLCTP
	DeepRiPP	Lanthipeptide class III/IV	MALLDLQAMDTPQEEAVGDLATGSQISLLICEYSSLSVTLCTP
	PRISM	Lanthipeptide class III/IV	MALLDLQAMDTPQEEAVGDLA TGSQISLLICEYSSLSVTLCTP
	RiPPMiner	Lanthipeptide class III/IV	MKTDLSWVLNDVLEVRGARHAILVSGDGLLLQSSDGIERGEAET NAAAMSSMQSLSRVASFVGLGRGVWKQTLMEYDGGWIFLIA AGQGAYLAVSAAVDVDMESMSIRMQKTVASLSRAMGVAPR S NNGVSV
3	antiSMASH	Not identified	
	BAGEL	Not identified	
	DeepRiPP	Not identified	
	PRISM	Not identified	
	RiPPMiner	Linaridin	MTMTPFALQYARPQTGEPAAPYAFDQAEQVNVLAGGGYAAED TALLARLGSTASTAGSKTHWDD
4	antiSMASH	Not identified	
	BAGEL	Linocin M18-like	
	DeepRiPP	Not identified	
	PRISM	Not identified	
	RiPPMiner	Not identified	
5	antiSMASH	Lanthipeptide	
	BAGEL	Lanthipeptide class IV	
	DeepRiPP	Lanthipeptide class III/IV	MSENTSPETPETVEAPEVEAHSASVLDLQGTTSDQEH IADGNC ISVLSVENQK
	PRISM	Lanthipeptide	VTSGTRPVKSWRIQLSSVGSFITCSVRCCRRRWSDRGGSCSCVTS GTRPVKSWRIQLSSVGSFITCSVRCCRRRWSDRGGSCS
	RiPPMiner	Lanthipeptide class II/III/IV	MQIDDAATHLTDGAGESAGA HPVGAITLFTTGALGLRSRLLSAS EGGASYSHELPWTTMTAPQ

Figure 16: Genome mining outputs from analysis of the *Streptomyces albus* J1074 genome. Predicted core peptides are highlighted in green.

2.3.2. Using RiPPER to Identify Novel RiPP Gene Clusters Associated with YcaO-domain Proteins

As discussed previously, it is known that RiPP gene clusters can be difficult to detect by genome mining, and the comparative analysis of tools described above highlights that not every available tool identifies every RiPP gene cluster. Following the success of the TfuA-led genome mining with RiPPER, we were interested to assess further unexplored diversity of RiPP gene clusters associated with YcaO-domain proteins. Although YcaO-domain proteins have several characterised catalytic roles including azoline formation, thioamide formation and amidine formation (116,204,205), the genetic and chemical context of many of these enzymes still remains uncharacterised, representing a large untapped source of potentially novel biochemical diversity. In order to investigate this, we decided to use Actinobacterial YcaO-domain proteins as 'bait' sequences for a RiPPER analysis.

2.3.2.1. Analysis of Actinobacterial YcaO-domain proteins

Using CDART (Conserved Domain Architecture Retrieval Tool) (301) on NCBI Genbank, I identified over 9,000 standalone proteins (i.e. those not fused to an additional domain) containing YcaO-domains. These were present predominantly in bacteria and archaea, with some also identified in fungi, plants and animals. For the focus of this study, I filtered these proteins down to those present in Actinobacteria, of which there were 2,574 sequences. Proteins fewer than 350 amino acids in length were excluded so that the results were not skewed by incomplete or truncated sequences. The remaining 2,338 sequences were then analysed with EFI-EST (Enzyme Similarity Tool) (302), applying a 95% maximum identify cut off to account for sequence duplications. The resulting sequence similarity network (SNN) obtained from EFI-EST contained 1,514 proteins, which grouped into various sub-networks of related YcaO-domain proteins (**Figure 17**).

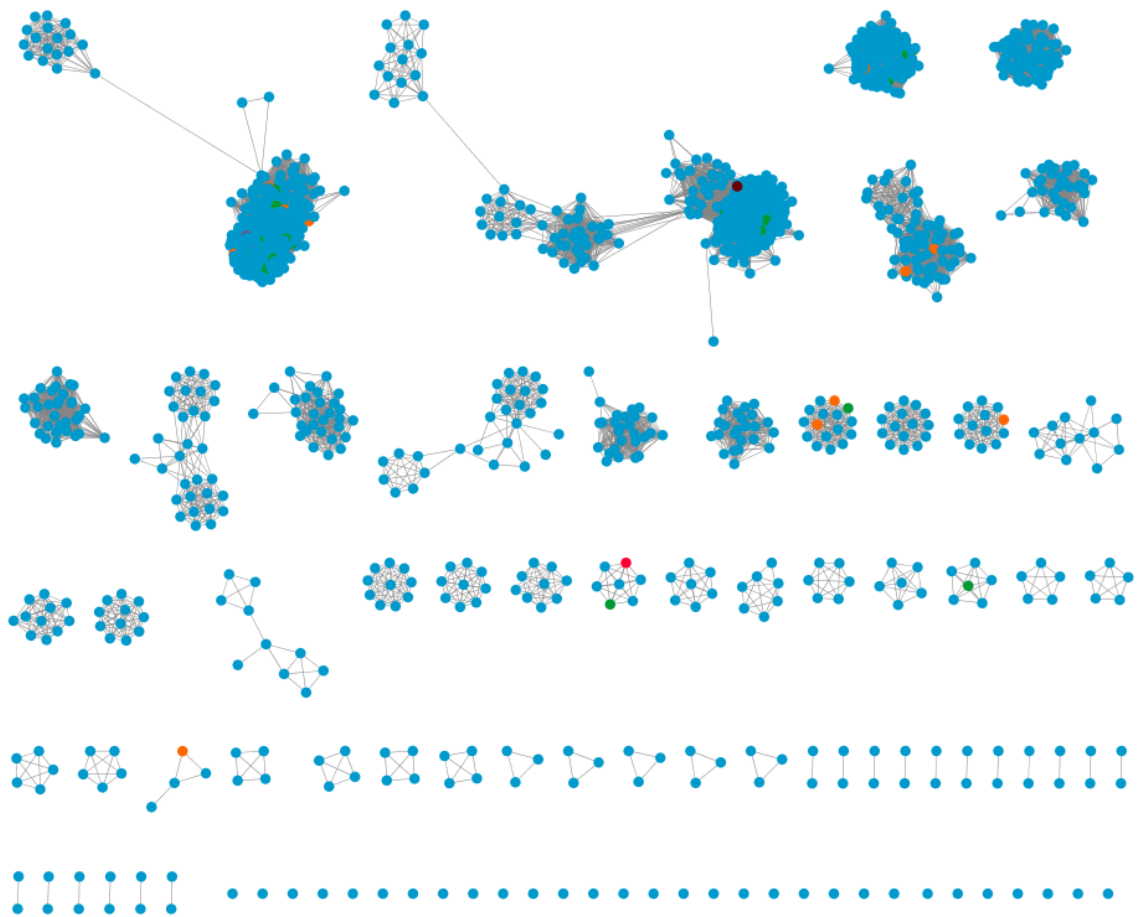


Figure 17: EFI-EST analysis of Actinobacterial YcaO-domain proteins. Image shows sequence similarity network (SSN) of 1,514 Actinobacterial YcaO-domain proteins retrieved from CDART. Blue nodes represent five or fewer protein IDs, green nodes represent 6-10 protein IDs, orange nodes represent 11-18 protein IDs, dark pink nodes represent 22-41 protein IDs, and the dark red node represents 92 protein IDs.

2.3.2.2. Retrieval and analysis of precursor peptides associated with YcaO-domain proteins

In order to identify some of the biosynthetic gene clusters associated with the uncharacterised Actinobacterial YcaO proteins, I first wanted to identify the associated precursor peptides. To achieve this, a RiPPER (124) analysis was carried out by Andrew Truman and Govind Chandra using the previously obtained 1,514 protein sequences as the input (methods described in section 6.3.2.1.). In total, 2,492 putative precursor peptides were retrieved, corresponding to short peptides within 6 kb of each YcaO-domain protein analysed. These peptides were subjected to an EGN (298) analysis carried out by Andrew Truman, with the resulting peptide networks shown in **Figure 18**. A 40% identity cut-off was used for creation of networks, as this resulted in clustering of known RiPP subfamilies. I annotated these networks based on the NCBI HMM domains associated with each precursor peptide. This showed that the precursor peptides within network 8 display homology to thiocillins, as do the majority of precursors within network 2. Precursor networks 11 and 13 show homology to thiazolylpeptides and network 12 shows homology to bottromycins. A subfamily of network 6 shows homology to thioamitides. Interestingly however, many of the precursor peptide networks did not show homology to known RiPP families, indicating that the RiPPER analysis had uncovered several previously unknown RiPP subfamilies. Several studies have highlighted expansion of the known biosynthetic landscape of RiPPs (252,253,303,304), but these data obtained from RiPPER contribute even further RiPP biosynthetic diversity that has not been previously reported. After submitting representative precursor sequences from networks 3, 4, 5, 7, 9 and 10 to RiPPMiner (281) and DeepRiPP NLPPrecursor (290), the majority of these precursors were predicted as “non-RiPP”. Despite this, the genetic context of these precursor peptides suggest that they are associated with genuine RiPP biosynthetic pathways, with the co-occurrence of putative RiPP tailoring genes within an operon. Therefore, it is likely that the prediction tools are not optimised for identification of all types of RiPP precursor, especially those that might represent novel subfamilies.

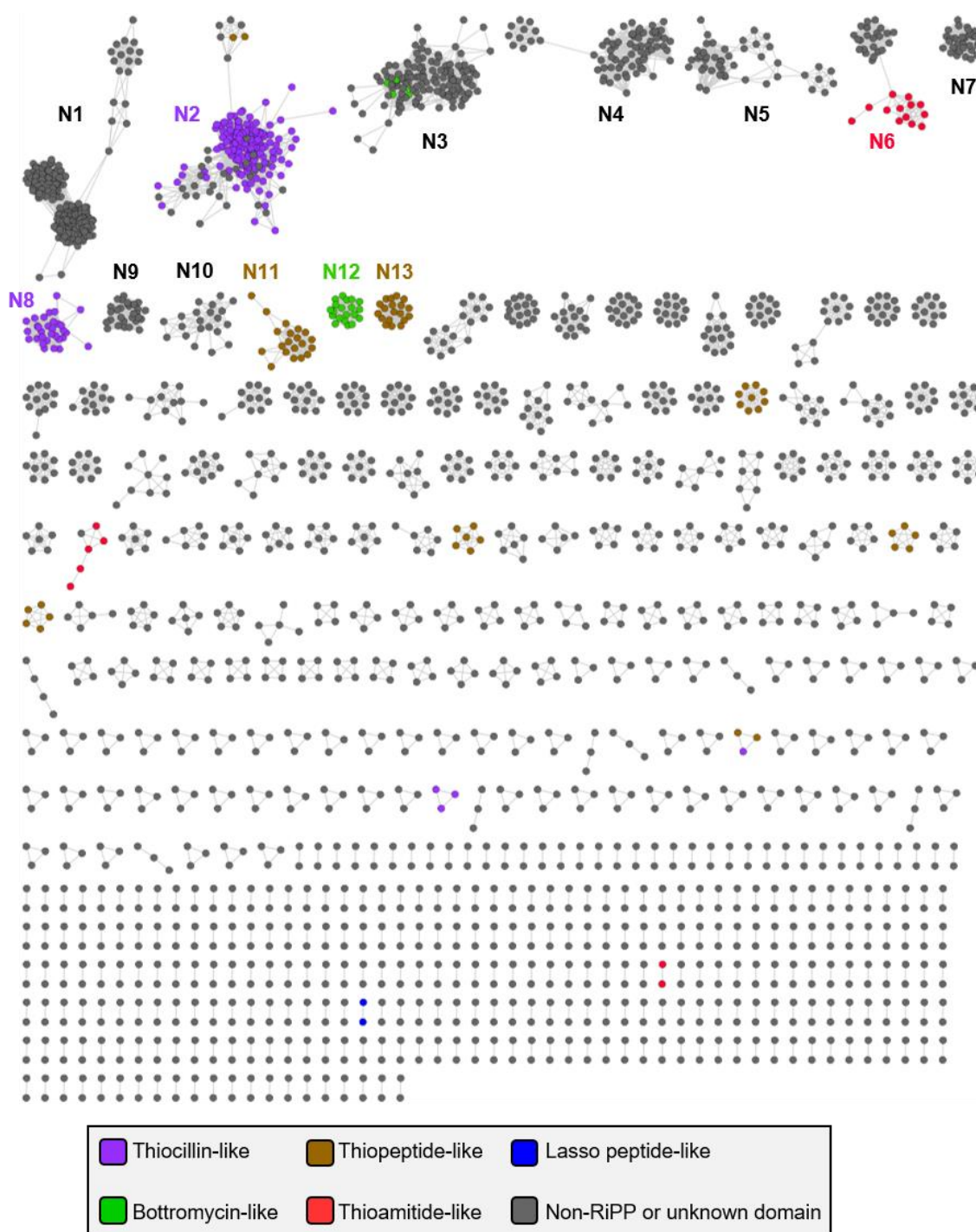


Figure 18: Precursor peptide networks identified from RiPPER analysis of Actinobacterial YcaO-domain proteins. Sequences were analysed by EGN using a minimum sequence identity cut-off of 40% and visualised on Cytoscape. This percentage identity cut-off was chosen as it resulted in network clustering of known RiPP precursor families. Peptides associated with known RiPP families are highlighted: green= bottromycin family (NCBI HMM domain NF033414), brown= thiopeptide family (NF033400 and NF033399), red= thioamide family (NF033415), purple= thiocillin-like family (NF033482 and NF033401), dark blue= lasso peptide family (NF033521).

2.3.2.3. Phylogenetic analysis of associated YcaO-domain proteins

The 1,514 YcaO-domain protein sequences associated with the identified precursor peptides were aligned using MUSCLE (264) followed by construction of a phylogenetic tree using RAxML on the CIPRES Science Gateway (305) (methods described in section 6.3.1.2.). The tree was then annotated based on the presence of known NCBI HMM domains in the precursor peptides identified by RiPPER. This annotation shows that YcaO proteins involved in biosynthesis of certain RiPP families are clustered within subclades of the phylogenetic tree (**Figure 19**). This includes YcaO proteins involved in the biosynthesis of bottromycins (two separate clades for BtmE-like and BtmF-like proteins), thioamitides, thiocillin-like metabolites and thiopeptides. Notably however, the biosynthetic potential of the majority of these YcaO proteins is unknown, based on the absence of known HMM domains. These enzymes might therefore carry out potentially rare or novel biochemical transformations.

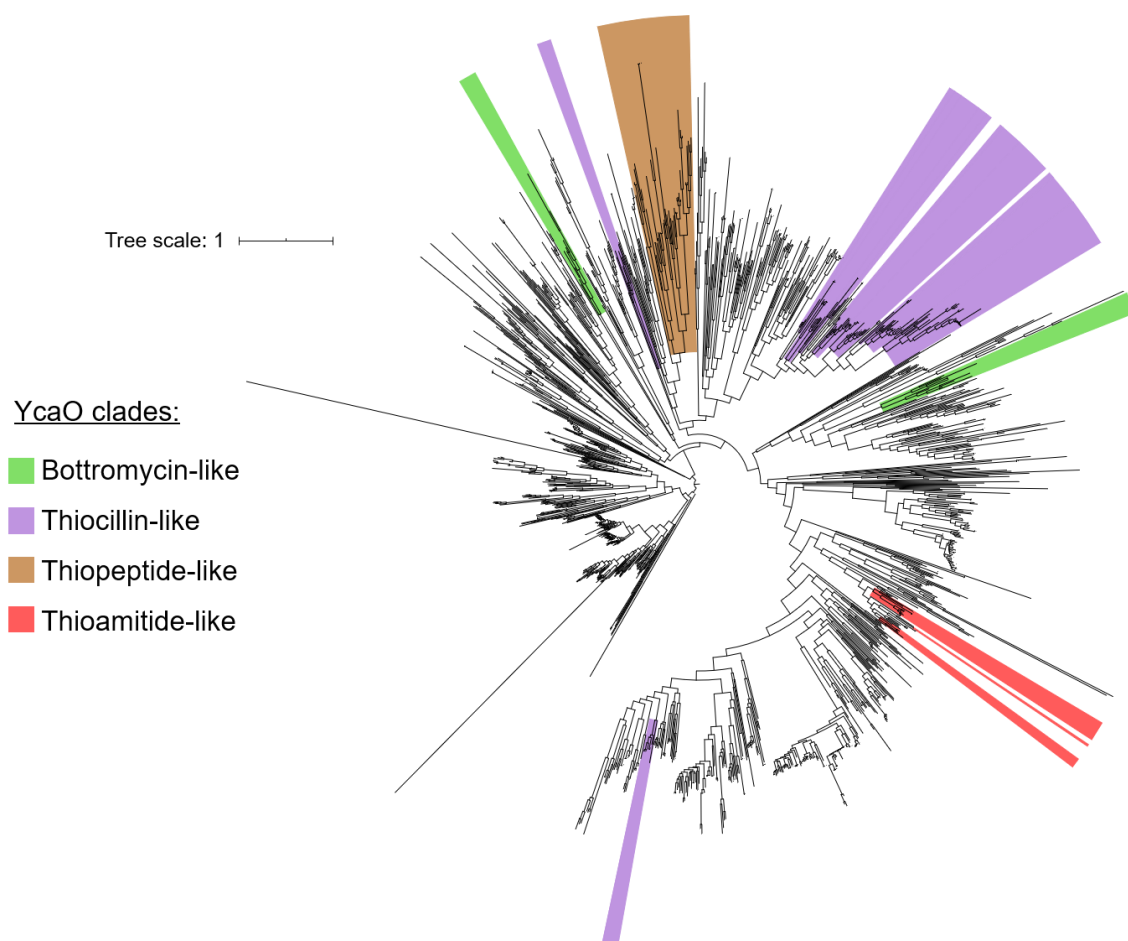


Figure 19: Phylogenetic tree of standalone YcaO-domain proteins from Actinobacteria. Proteins with homology to known RiPP biosynthetic classes are highlighted: green= bottromycin family (NCBI HMM domain NF033414), brown= thiopeptide family (NF033400 and NF033399), red= thioamitide family (NF033415) and purple= thiocillin-like family (NF033482 and NF033401).

2.3.3. Bioinformatic Analysis of a Novel RiPP Precursor Peptide Family

The largest precursor peptide network identified by RiPPER (network 1) is the focus of this thesis. This network represents 260 precursor peptide sequences associated with Actinobacterial YcaO-domain proteins. Like many of the other peptides identified with RiPPER, these sequences do not show homology to known RiPP classes, and DeepRiPP and RiPPMiner predicted the majority of sequences to be non-RiPP. The fact that these peptides are the most abundant precursors associated with standalone Actinobacterial YcaO proteins, and yet do not appear to belong to known RiPP families makes these an intriguing group of RiPPs to study. The precursor sequences within this network are therefore studied in further detail in the following sections.

2.3.3.1. Sequence alignment and motif identification

Out of the 260 precursor peptides in network 1, 29 sequences were duplications due to the presence of more than one YcaO protein in the associated gene clusters. These sequences were therefore removed prior to further analysis. The remaining 231 precursor sequences varied greatly in length, between 31 and 89 amino acid residues, indicating diversity within this network of peptides. The sequences were aligned using MUSCLE (264) (**Appendix Figure 83**), which highlighted that an 'ALV' (alanine-leucine-valine) motif was conserved across 216/231 sequences. This suggested that these peptides are related and are likely to be genuine precursors from a large novel RiPP family. Interestingly, some of these peptides contained two or three repetitions of the ALV motif, which could suggest that multiple metabolites might be produced by a single precursor peptide. This was observed with the thiovarsolins, whose precursor peptides contained repetitive 'APR' motifs (124). To investigate precursor motifs further, all 231 sequences were subjected to a MEME analysis (306), searching for up to five motifs within the sequences (**Appendix Figure 84**). This analysis also showed that ALV-containing motifs were present in the majority of sequences (**Figure 20**).



Figure 20: MEME analysis of peptide sequences from novel RiPP precursor network 1, identified following RiPPER analysis of Actinobacterial YcaO-domain proteins. Figure shows the top motif identified from the analysis, which highlights a conserved ALV motif.

Interestingly, manual analysis of the peptide sequences indicated that most of the peptides contained either a 'QGPQT'-like motif (motif A) or a 'HxSaxH'-like motif (motif B) adjacent to the conserved ALV. The distribution of these two motifs corresponds to the two peptide sub-networks within network 1 (**Figure 21**). I separately submitted the sequences within each of these sub-networks for a MEME analysis, and the top resulting motif in each case highlights this observation (**Figure 21**). These two groups of peptides might therefore produce related yet structurally distinct final metabolites. The other loosely connected nodes did not appear to have a similar conserved motif. These peptides might still represent genuine RiPPs that are less similar to the other peptides identified.

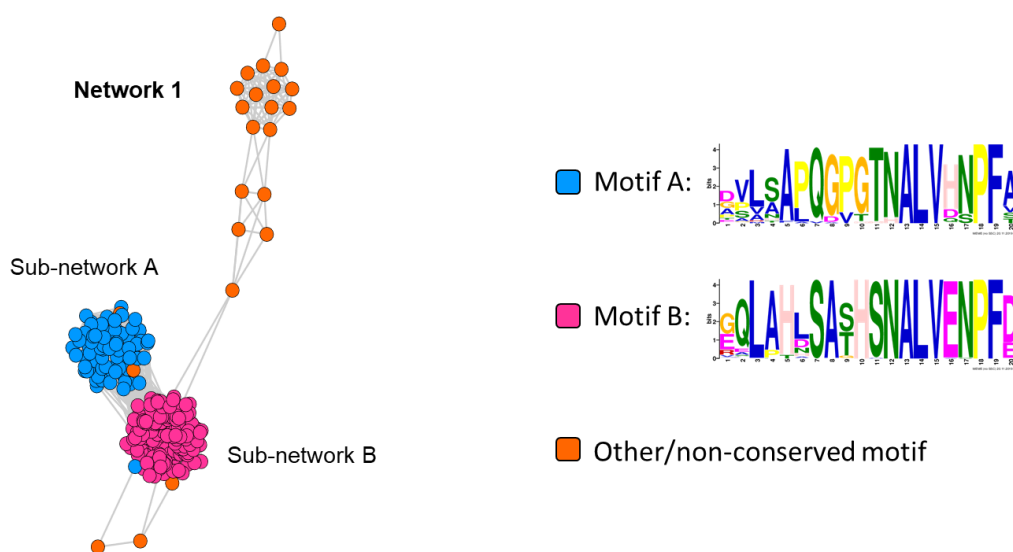


Figure 21: Comparison of precursor network 1 with corresponding sequence motifs. Precursor peptide sub-networks correspond to the presence of one of two conserved sequence motifs identified from MEME analysis of peptide sequences within each sub-network.

2.3.3.2. Distribution of identified precursor peptides in nature

Although all of the identified precursor peptides from network 1 were retrieved from Actinobacteria, a further BLAST analysis of the YcaO-domain proteins associated with these precursors revealed that six further related peptides are encoded in Firmicutes genomes: three from *Bacillus* and three from *Paenibacillus* species. These sequences were therefore included in all further analyses, bringing the total number of precursor peptides to 237. Overall, these precursor peptides and associated biosynthetic gene clusters are widespread in nature, present in two bacterial phyla, eight orders, 22 families and 57 different genera. The majority of the precursor peptides (86/237) were present in Streptomycetaceae genomes: 84 from *Streptomyces* species and two from *Kitasatospora* species. The second most abundant family were Microbacteriaceae (40/237 sequences), including the *Clavibacter* and *Microbacterium* genera. Other

bacterial families containing these precursor peptides included Pseudonocardiaceae (22/237 sequences), Nocardiaceae (20/237 sequences), Nocardiopsaceae (13/237 sequences) and Frankiaceae (12/237 sequences). Other bacterial families containing fewer than ten representative precursor peptides include Actinomycetaceae, Micrococcaceae, Bacillaceae, Dermatophilaceae, Streptosporangineae, Actinosynnemataceae, Nocardiodaceae, Cryptosporangiaceae, Promicromonosporaceae, Dermabacteraceae, Jiangellaceae, Glycomycetaceae, Geodermatophilaceae, Nakamurellaceae, Propionibacteriaceae and Thermomonosporaceae. The fact that these novel gene clusters are conserved across so many bacterial families suggests that the RiPPs being produced are playing an important biological function for the producing organisms. Interestingly, the species harbouring these gene clusters also occupy a range of ecological niches. This includes soil-dwelling streptomycetes and symbiotic nitrogen-fixing *Frankia* species that live in nodules of actinorhizal plants (307).

In order to analyse the phylogenetic relationship of these precursor peptides in more detail, a further EGN analysis was carried out using a lower sequence identity cut-off of 80%. The 19 resulting networks of peptide subfamilies are shown in **Figure 22**. Each of the nodes are colour-coded based on the bacterial family that the precursor peptide derives from in order to give a visual representation of the evolution and phylogenetic relationship of each peptide. The peptides containing either motif A or motif B are also indicated. This analysis shows that the majority of precursor peptides cluster into networks relating to the bacterial family they belong to, suggesting that these sequences might have co-evolved. However, there are also some examples of networks containing precursor peptides from a variety of different bacterial families, such as sub-family 2. This analysis also highlights that not all precursors within a given bacterial family have similar sequences. For example, sub-families one and five are distinct from each other whilst both representing Streptomycetaceae species. Similarly, peptides from Microbacteriaceae are present in sub-families two, four, nine, 11 and 13. This could be due to divergent evolution of some precursor sequences, or horizontal gene transfer.

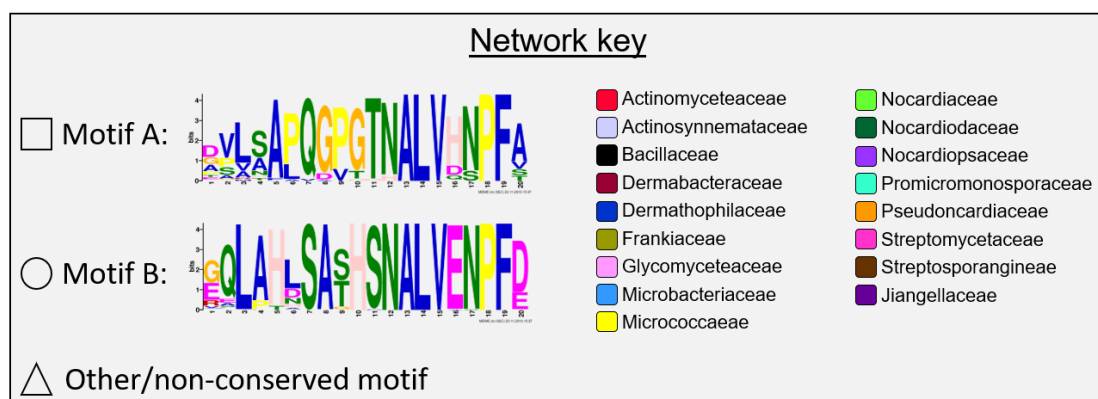
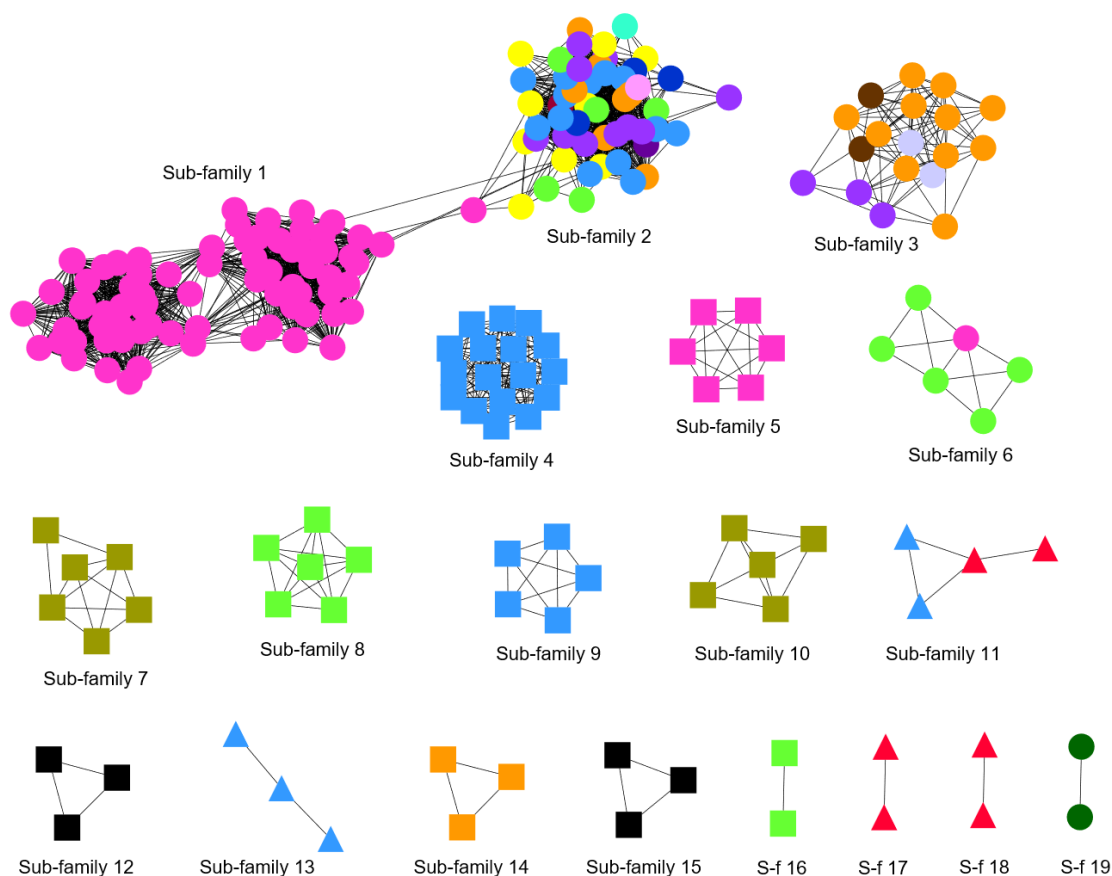


Figure 22: EGN networking analysis of precursor peptides within network 1 (shown in Figure 18). Networks analysed using a lower sequence identify cut-off of 80%. Nodes are colour-coded by bacterial family and the occurrence of different sequence motifs is indicated by node shape.

2.3.3.3. Relationship between newly identified precursor peptides and their corresponding YcaO-domain proteins

Following the evolutionary analysis of the precursor peptides identified by RiPPER, I was interested to see how this related to the YcaO-domain proteins that are present in the associated biosynthetic gene clusters of each of the precursors present in network 1. The sequences of all of the YcaO-domain proteins associated with the identified precursor peptides were therefore aligned using MUSCLE (264) and a phylogenetic tree was created using RAxML on the CIPRES Science Gateway (305) (methods described in section 6.3.1.2.). Each leaf is colour-coded by bacterial family, showing that the YcaO-domain proteins cluster in a similar way as the corresponding precursor peptide sub-networks (**Figure 23**).

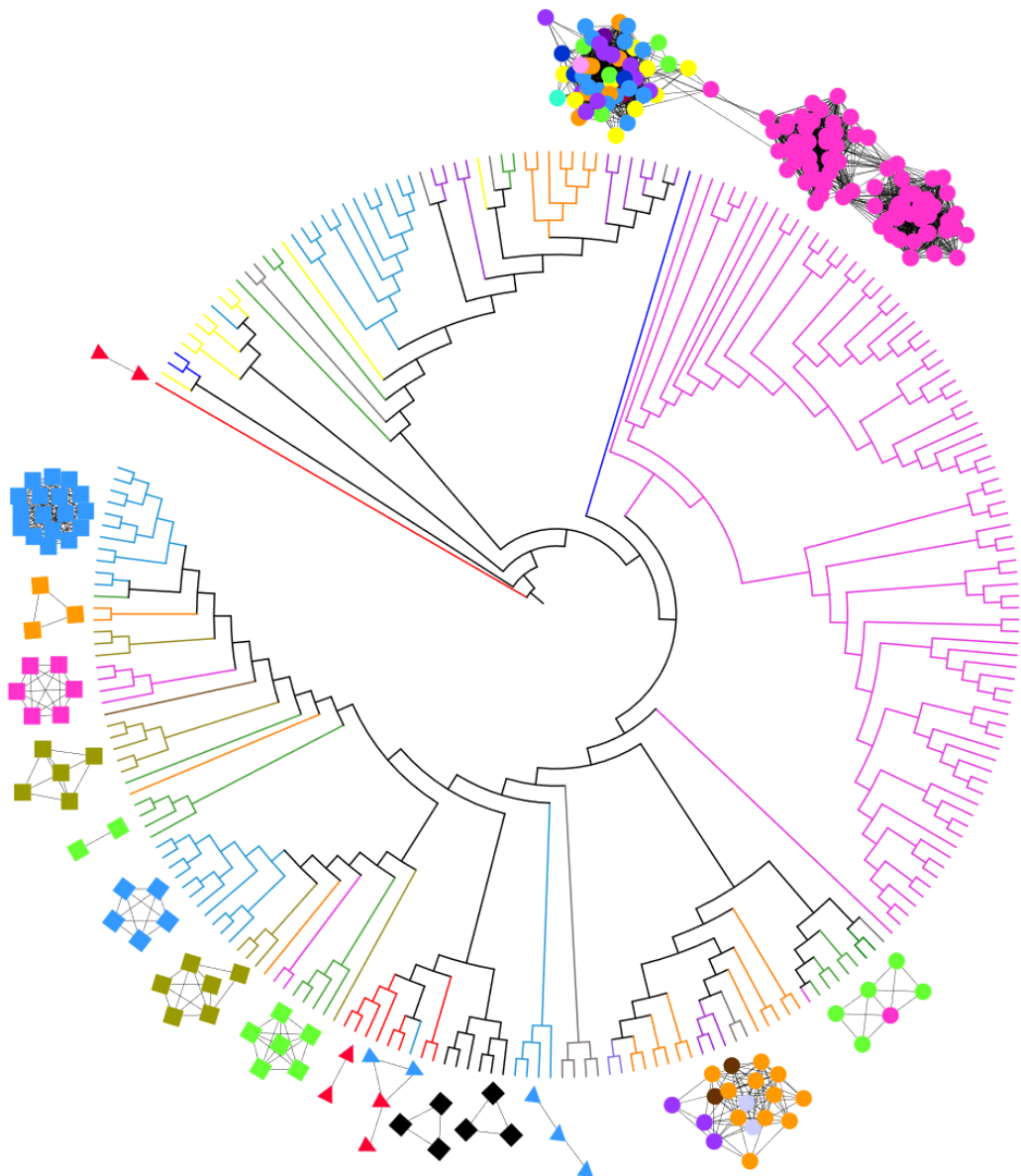
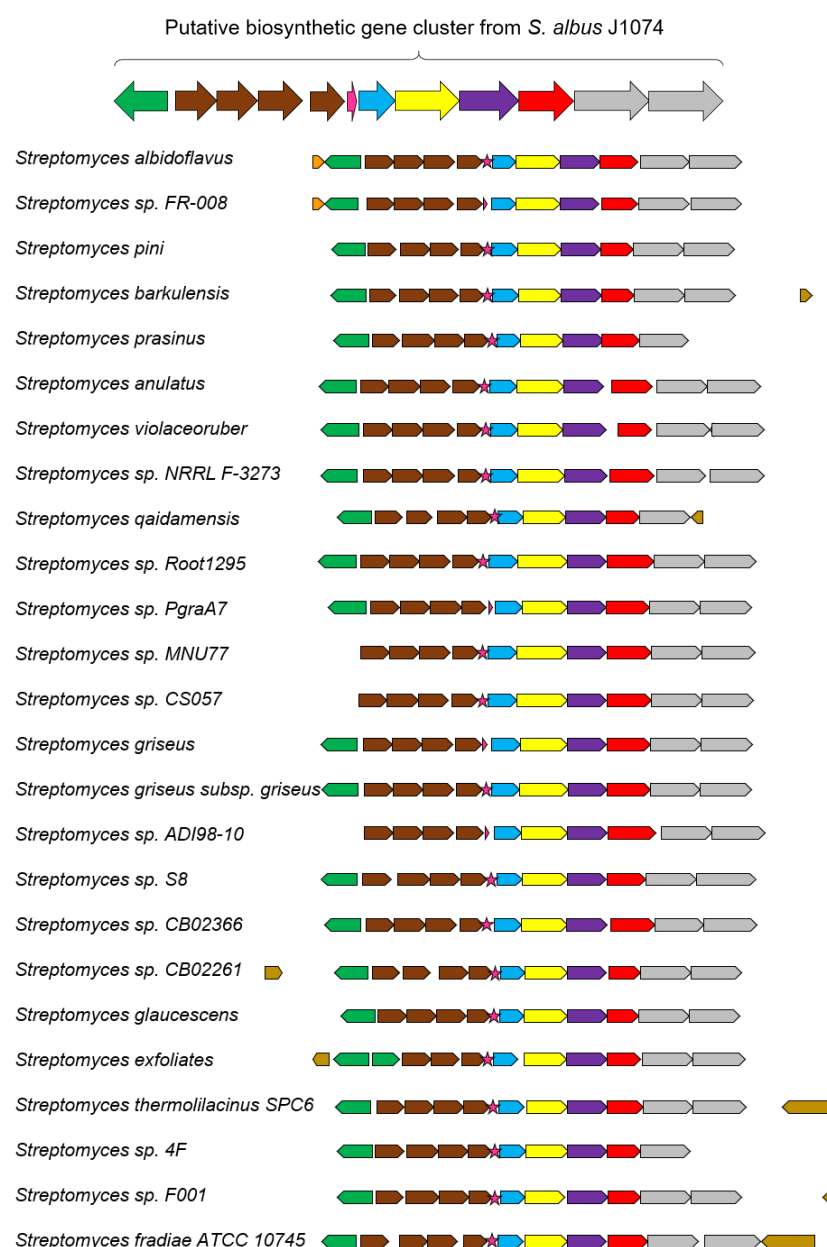


Figure 23: Phylogenetic tree of YcaO-domain proteins associated with precursor network 1. Precursor peptide networks from Figure 22 are mapped onto the tree highlighting the relationship between YcaO phylogeny and precursor peptide evolution. Nodes are colour-coded by bacterial family (as in Figure 22).

2.3.4. Comparison of Gene Cluster Architectures

2.3.4.1. MultiGeneBlast analysis

In order to put the identified precursor peptides and corresponding YcaO proteins in the context of full biosynthetic gene clusters, a MultiGeneBlast (308) analysis was carried out by Andrew Truman. For the input, all genes that were likely to form the novel RiPP biosynthetic gene cluster from *S. albus* J1074 were used. This included a hypothetical protein homologous to an oxidoreductase, four iron transporter genes with homology to the FecBCDE system, the putative precursor peptide, a conserved hypothetical protein homologous to an E1-ubiquitin activating enzyme, a hydrolase, the YcaO-domain protein, a flavin-dependent dehydrogenase, and two ABC transporters. The first ~90 gene clusters retrieved by MultiGeneBlast show close homology to the *S. albus* J1074 gene cluster, suggestive of a large family of related RiPPs (**Figure 24**).



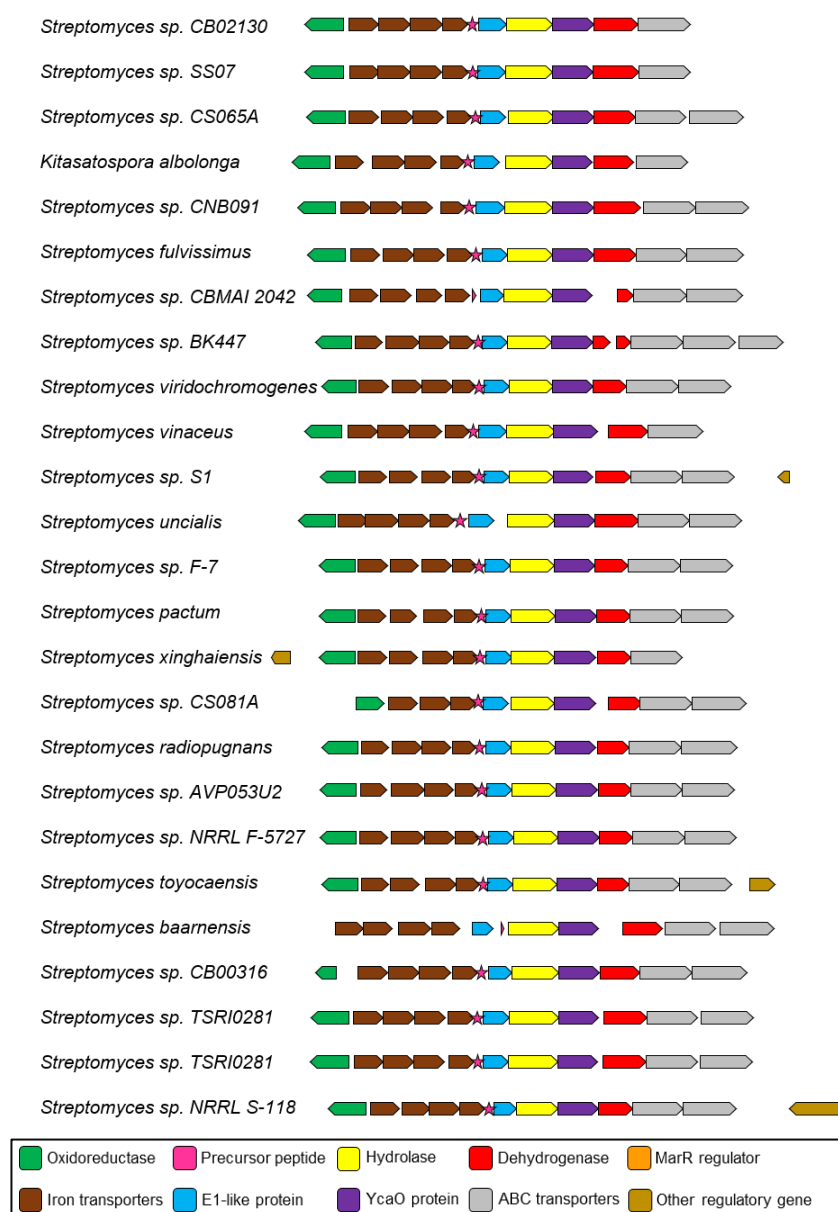
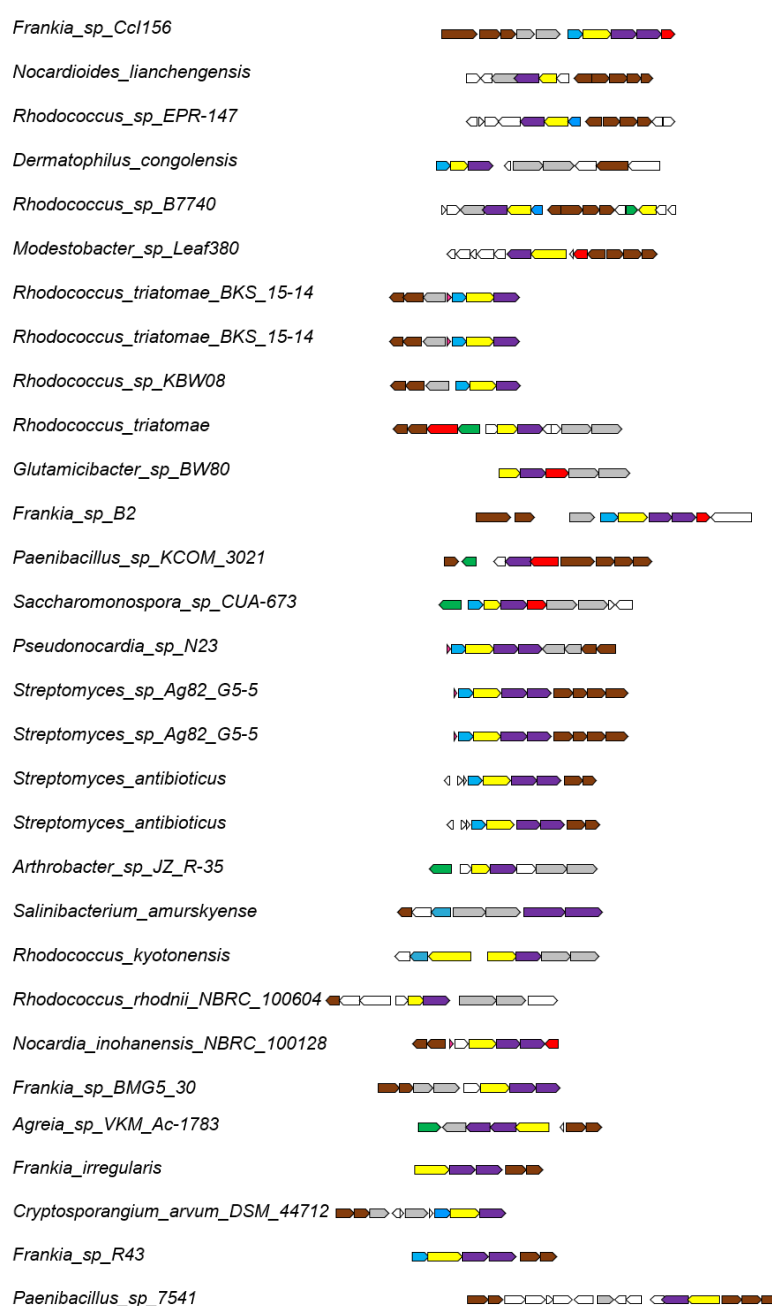


Figure 24: Example BGCs homologous to that of *S. albus* J1074. Obtained from MultiGeneBlast analysis using the *S. albus* J1074 pathway as input. Precursor peptides are indicated by pink stars, other biosynthetic genes retrieved by MultiGeneBlast are indicated by coloured arrows.

The majority of precursor peptides retrieved by RiPPER were not annotated in the MultiGeneBlast output. Interestingly, a further analysis by Govind Chandra revealed that only 78 of the identified precursor peptides (33%) were originally annotated in genomes, demonstrating that this novel genome mining approach can successfully detect and annotate previously unknown RiPP precursor peptides. Beyond the biosynthetic gene clusters showing close homology to the *S. albus* J1074 pathway, numerous additional gene clusters display further biosynthetic diversity (**Figure 25**). For example, several gene clusters contain two YcaO-domain proteins, indicating that the resulting metabolites might contain different or additional post-translational modifications. Many gene clusters also lack the E1-like homolog, the dehydrogenase and the oxidoreductase genes.



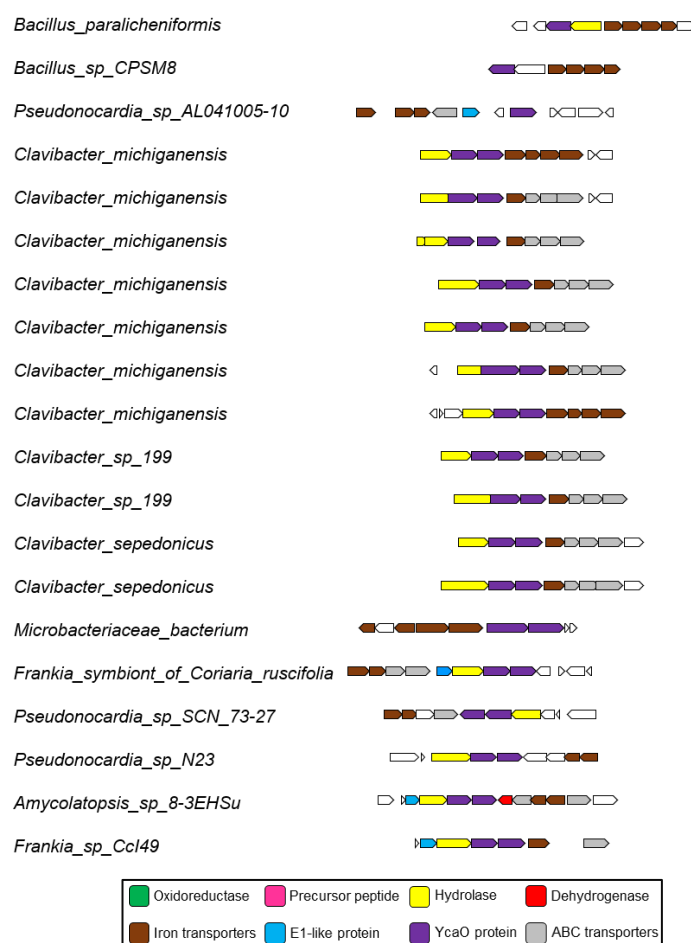


Figure 25: Example biosynthetic gene clusters that are related to that of *S. albus* J1074. Obtained from MultiGeneBlast analysis using the *S. albus* J1074 pathway as input. Pathways contain the conserved YcaO-domain protein but show genetic diversity in comparison to the input gene cluster.

2.3.4.2. Pathway regulation

In the model *S. albus* J1074 gene cluster, a MarR regulator is encoded next to the proposed minimal biosynthetic gene cluster, which could plausibly be involved in regulation of this RiPP. However, despite the fact that the core biosynthetic genes are conserved across approximately 90 pathways, this MarR gene is only conserved in one other gene cluster, from *Streptomyces* sp. FR-008, according to the MultiGeneBlast output. Some other pathways contain regulatory genes within the vicinity of the core gene cluster, but these vary in position and nature. For example, some of these regulators are annotated as CynR-like, DeoR-like, TetR-like or LuxR-like. This suggests that the regulation of these gene clusters might greatly vary, despite the conserved nature of other biosynthetic genes.

2.3.4.3. Comparison of gene cluster architecture with the associated YcaO and precursor phylogeny

Based on the observation of gene cluster diversity, I was interested to assess whether this corresponded to the different precursor peptide sub-networks. I manually compared precursor peptides containing motif A or B to their corresponding gene cluster, and then compared this to the location of corresponding YcaO-domain proteins on the Actinobacterial YcaO phylogenetic tree. These comparisons are shown in **Figure 26**.

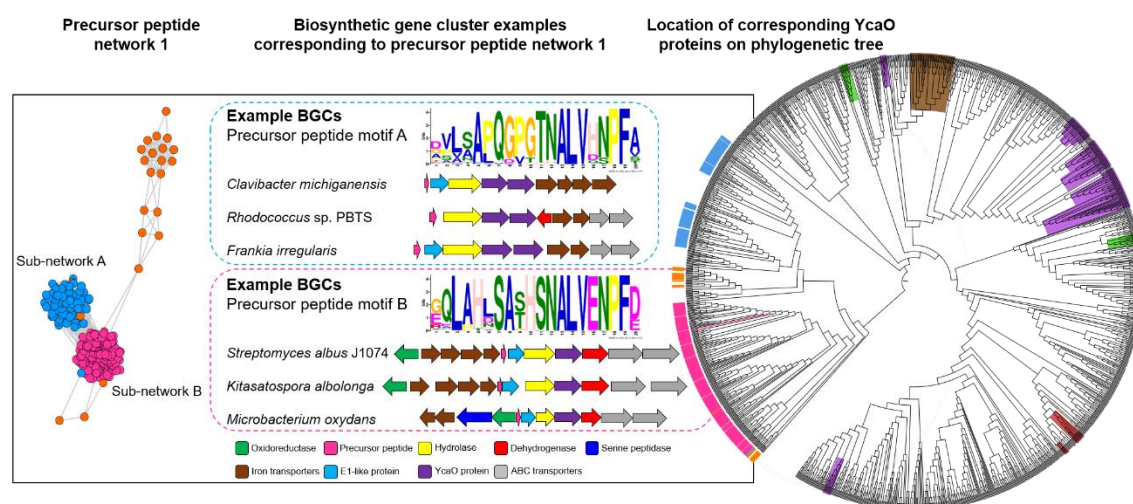


Figure 26: Comparison of YcaO phylogeny, gene clusters and precursor sub-networks. Actinobacterial YcaO phylogenetic tree (as in Figure 19, with branch lengths not shown) with examples of corresponding biosynthetic gene clusters and their relationship to precursor peptide sub-networks (from Figure 21).

From this analysis, it is clear that there is a direct link between the precursor peptide sequence, genetic architecture and location of the corresponding YcaO protein on the phylogenetic tree. Precursor peptides containing sequence motif B are present in gene clusters similar to that of *S. albus* J1074, containing one YcaO-domain protein. Precursor peptides containing sequence motif A are present in a second distinct group of biosynthetic genes clusters, which typically contain two YcaO-domain proteins. The YcaO-domain proteins associated with these two groups of gene cluster also tightly map to two distinct subclades on the YcaO phylogenetic tree. Taking these results together, it is likely that in the majority of cases the newly identified precursor peptides and YcaO-domain proteins co-evolved together within their biosynthetic gene clusters. Overall, these data represent a large amount of biochemical diversity among a widespread novel family of RiPP biosynthetic gene clusters.

2.4. Chapter Summary

This chapter describes a range of different genome mining tools for RiPPs and demonstrated how these can be applied to the discovery of novel RiPP precursor peptides and associated biosynthetic gene clusters. The key finding from this chapter was the discovery of novel families of RiPP precursor peptides associated with standalone YcaO-domain proteins from Actinobacteria, using the recently developed RiPPER tool. The majority of these precursor peptides were previously uncharacterised, with no homology to known RiPP classes. The largest precursor peptide network retrieved by RiPPER was analysed in closer detail through motif analysis and biosynthetic gene cluster comparison. The widespread nature of these novel RiPP pathways as well as their sequence and genetic diversity make these gene clusters an intriguing point of focus for this thesis. A comparative analysis of genome mining tools was also carried out using the whole genome of *Streptomyces albus* J1074, a model streptomycete that contains one of the newly identified RiPP biosynthetic gene clusters. This demonstrated that different genome mining tools vary in their ability to identify particular RiPP gene clusters and precursor peptides. Furthermore, none of the tools identified the novel RiPP gene cluster discovered from our RiPPER analysis. Overall, these analyses highlight that many currently available genome mining tools have overlooked untapped RiPP diversity, and the use of more targeted gene-led tools such as RiPPER is a valuable approach for identifying true RiPP novelty.

The following chapter will focus on the characterisation of one of the novel RiPP biosynthetic gene clusters identified.

Chapter 3: Cloning and Expression of a Novel RiPP Gene Cluster from *Streptomyces albus* J1074

3.1. Introduction

Following the genome mining experiments described in Chapter 2, this chapter will focus on the characterisation of one of the novel RiPP gene clusters identified with RiPPER, using a combination of cloning, genetic experiments and metabolomic screening.

3.1.1. Genetic Manipulation of Natural Product Biosynthetic Gene Clusters

The genetic study and manipulation of microorganisms has historically been hindered by the variable ability of microbes to uptake foreign plasmids, but recent advances in genetic engineering have made the direct study of microbial gene clusters increasingly more achievable. The development of a range of genome editing tools has allowed for manipulation of biosynthetic gene clusters *in situ*, and molecular techniques based on homologous recombination and nuclease activity have allowed for efficient deletion and disruption of biosynthetic genes. Several recombination-based techniques have been developed based on the lambda (λ) Red system in *E. coli* (309–311), which are often referred to as recombineering (recombination-mediated genetic engineering) approaches. This includes techniques such as PCR-targeting that was first developed for the deletion of genes from the *E. coli* K-12 chromosome (312) and later adapted for *Streptomyces* species with development of the ReDirect method (309). Early nuclease-based editing included the use of meganucleases such as 18 bp cutter I-SceI (313) to introduce double strand breaks at target DNA sites. Later approaches explored the use of zinc finger nucleases (314) and transcription activator-like effector nucleases (TALENs) (315). More recently, advances in CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat DNA sequences) technology have revolutionised genome editing in both eukaryotes and prokaryotes. CRISPR-associated (Cas) endonuclease genes such as Cas9 can be programmed to target specific DNA sequences in bacteria using custom made guide RNAs (319,320). CRISPR has also been optimised for use in *Streptomyces* species with the construction of the pCRISPomyces plasmids in 2015 (317). Many molecular microbiology approaches also rely on phage integrases for the expression of genes. These enzymes mediate site-specific recombination between two DNA recognition sequences, and include the tyrosine integrases such as the *E. coli* λ

phage (318), and the serine integrases such as the ϕ C31 phage from *Streptomyces lividans* (319) and the ϕ BT1 phage from *S. coelicolor* (320).

3.1.2. Cloning Natural Product Biosynthetic Gene Clusters

Whilst genetic engineering approaches allow the study of biosynthetic gene clusters in their native context, another common approach in natural products research is to clone a whole biosynthetic gene cluster for heterologous expression. This can be useful if the native producer of a gene cluster is difficult to directly genetically manipulate. Whilst PCR amplification of large DNA fragments is becoming increasingly more accurate, it can be difficult to clone large biosynthetic gene clusters, and many pathways for nonribosomal peptides and polyketides contain highly repetitive DNA regions that are challenging for PCR amplification (321). Recently, a direct pathway cloning (DiPaC) approach was described (322), which relies on long-range PCR and *in vitro* DNA assembly. Another cloning method is transformation-associated recombination (TAR) cloning, which takes advantage of the high levels of homologous recombination that occurs between the ends of DNA sequences in yeast cells during transformation (323). TAR cloning was first developed for the cloning of mammalian DNA (324,325), and later adapted for use in prokaryotes by integrating an autonomously replicating sequence (*ARS*) element into the TAR vector to allow it to act as an artificial chromosome (326). Another method for cluster cloning is Cas9-assisted targeting of chromosome segments (CATCH), which relies on an RNA-guided Cas9 nuclease which is directed to cut at two designated loci in the target genome (327).

3.1.3. *Streptomyces albus* J1074 as a Model Organism

The genome mining analyses described in Chapter 2 identified over 230 novel RiPP biosynthetic gene clusters present in a wide range of Actinobacteria and Firmicutes. As an exemplar pathway for this study, I chose to investigate the gene cluster from *Streptomyces albus* J1074 (formally reclassified as *Streptomyces albidoflavus* J1074 (328)). The genetic architecture of the cluster from *S. albus* J1074 represents one of the most common pathways of those identified, with over 90 gene clusters with identical architectures identified by MultiGeneBlast (**Chapter 2, Figure 24**). Another reason *S. albus* J1074 was chosen for study is that it is a widely used model streptomycete that is known to be genetically tractable, and has a relatively fast growth rate for *Streptomyces* bacteria (329). The favourable growth traits of *S. albus* J1074 are partly due to its naturally minimised genome, owing to a low number of gene and operon duplicates (299). *S. albus* J1074 has one of the smallest known genomes for streptomycetes at 6,841,649 bp, which codes for 5,832 genes. The chromosome contains a large central core region and two small arms that are 0.3 Mb (left) and 0.4 Mb (right) in length (299).

3.2. Chapter Aims

The RiPP biosynthetic gene clusters identified from genome mining with RiPPER (Chapter 2) are hypothesised to represent novel RiPP families. The overall aim of this chapter was to investigate the metabolite(s) produced by an exemplar biosynthetic gene cluster from *S. albus* J1074, which represents one of a large novel RiPP family. To achieve this, the specific objectives of this chapter were:

- (i) TAR clone the model biosynthetic gene cluster from *S. albus* J1074
- (ii) Express the cloned gene cluster in a heterologous host for metabolomic analyses
- (iii) Construct a pathway mutant in *S. albus* J1074 to investigate metabolomic profiles of the native host
- (iv) Construct gene deletions within the cloned pathway to investigate roles of individual biosynthetic genes

3.3. Results and Discussion

3.3.1. Overview of the *Streptomyces albus* J1074 Biosynthetic Gene Cluster

The putative biosynthetic genes within the *S. albus* J1074 biosynthetic gene cluster (*ami*) were analysed using BLAST (266), CDART (301) and Phyre2 (267) to determine their likely enzymatic functions. A minimal biosynthetic gene cluster was also proposed, based on the genes that were conserved across several pathway homologues (indicated from the MultiGeneBlast analysis shown in section 2.3.4.1.) (**Figure 27**). This proposed minimal cluster comprises a putative oxidoreductase, a four-component iron transporter system, a precursor peptide, an E1-ubiquitin-like enzyme, a hydrolase, the YcaO-domain protein, a flavin-dependent dehydrogenase and two ABC transporters. Details of these genes are summarised in **Table 1**. The tight clustering of genes encoding the putative precursor peptide, E1-like protein, YcaO-domain protein and dehydrogenase is typical of gene clusters for linear azoline-containing peptides (LAPs) such as microcin B17 (330) and klebsazolicin (205). The hydrolase enzyme encoded in between the E1 and YcaO proteins could theoretically act as a peptidase that cleaves the modified core peptide from the leader and/or follower regions. The two transport systems could represent a coordinated import and export mechanism for the associated natural product, or for elements required for biosynthesis of the RiPP metabolite. The iron transport system could also indicate an important role for iron or other metals during biosynthesis of the associated RiPP, or could suggest a potential siderophore-like function of the metabolite. The gene encoding the oxidoreductase was indicated from MultiGeneBlast analysis to be conserved across gene clusters similar to the *S. albus* J1074 pathway (section 2.3.4.1. MultiGeneBlast analysis.), therefore this enzyme could play an important biosynthetic role. However, oxidoreductases are not typically seen in LAP gene clusters. Finally, the MarR regulator could be involved in regulation of the pathway, but this regulator was not conserved across similar biosynthetic gene clusters, suggesting that it may not be essential.

In *S. albus* J1074, this novel RiPP gene cluster is located in the small right-hand arm of the chromosome. According to the genome mining outputs from the *S. albus* J1074 genome (described in section 2.3.1.) the gene cluster is flanked tightly by two other metabolic pathways: a class IV lanthipeptide and a linocin M18-like bacteriocin. This is typical of secondary metabolite biosynthetic pathways that are often clustered together on a chromosome.

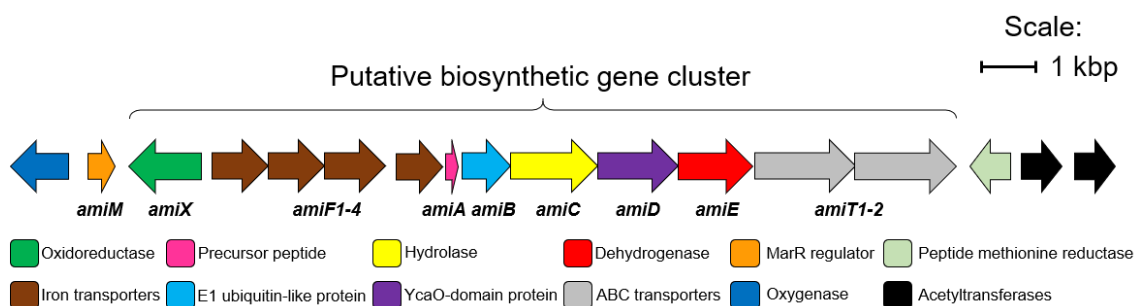


Figure 27: Putative RiPP biosynthetic gene cluster (*ami*) from *S. albus* J1074. Key genes proposed to be involved in biosynthesis of the final pathway-related metabolite are indicated.

Table 1: Details of predicted protein functions of the genes present in the *Streptomyces albus* J1074 BGC.

Gene name	Protein accession number	Pfam domain	Predicted function	Size (AA)	Size (kDa)
<i>amiM</i>	WP_085479280.1	No Pfam match	MarR regulator	147	15.9
<i>amiX</i>	WP_085479279.1	No Pfam match	Oxidoreductase	417	43.35
<i>amiF1</i>	WP_085479278.1	PF01497	Iron transporter (substrate-binding protein)	329	34.74
<i>amiF2</i>	WP_085479277.1	PF01032	Iron transporter (permease)	344	34.78
<i>amiF3</i>	WP_085479276.1	PF01032	Iron transporter (permease)	351	34.6
<i>amiF4</i>	WP_085479275.1	PF00005	Iron transporter (ATP-binding protein)	278	29.9
<i>amiA</i>	AMM12575.1	No Pfam match	Precursor peptide	44	4.6
<i>amiB</i>	WP_085479274.1	No Pfam match	E1-like protein	281	29.05
<i>amiC</i>	WP_129865485.1	PF02129	Hydrolase	508	54.26
<i>amiD</i>	WP_008409979.1	PF02624	YcaO-domain	455	48.6
<i>amiE</i>	WP_129857084.1	No Pfam match	Dehydrogenase	435	44.98
<i>amiT1</i>	WP_049977232.1	PF00005	ABC transporter (ATP-binding protein)	581	60.46
<i>amiT2</i>	WP_049977233.1	PF00005	ABC transporter (ATP binding protein)	594	60.7

3.3.2. TAR Cloning of the *S. albus* Biosynthetic Gene Cluster

The first aim of this chapter was to clone the model biosynthetic gene cluster from *S. albus* J1074. Although several methods for gene cluster cloning have been developed, such as PCR-based direct pathway cloning (DiPaC) (322) and Cas9-assisted targeting of chromosome segments (CATCH) (327), I chose TAR cloning as a method as it has previously been successfully used to capture large biosynthetic gene clusters from Actinobacteria, including by members of our lab group (124,151,331). The pCAP03 vector (332) was used as the backbone for capture of the gene cluster. pCAP03 is a shuttle vector containing elements that allow propagation and transfer in yeast cells, *E. coli* and *Streptomyces* species. These include a tryptophan auxotrophic marker (TRP1), an ARSH4/CEN6 region and the yeast URA3 gene encoding orotidine 5-phosphate decarboxylase (ODCase). This enzyme allows counter-selection using 5-FOA, which minimises false positive colonies that result from high levels of vector re-circularisation due to non-homologous end joining (326). The vector also contains a pUC *ori* for replication in *E. coli* and a kanamycin resistance gene (*aph(3)II*) for selection in *E. coli* and *Streptomyces* species. For conjugal transfer and chromosome integration in *Streptomyces* species, the vector contains an origin of transfer (*oriT*) and a ϕ C31 integrase and integration site.

3.3.2.1. TAR cloning design

As well as the set of biosynthetic genes proposed to comprise the target RiPP biosynthetic gene cluster, I designed my TAR cloning in such a way that some additional genes on either side of this cluster were also captured, including an oxygenase, peptide methionine reductase and two acetyltransferase genes (**Figure 27**). This was to ensure that all pathway elements required for biosynthesis of the target metabolite were incorporated into the expression vector. Restriction enzyme cut sites were identified (NsiI and SmlI) that would enable excision of this target genomic region (18.5 kbp) following digestion of *S. albus* J1074 genomic (g)DNA with these two enzymes. Regions of homology immediately adjacent to the cut sites were then used for design of the gene cluster capture vector to promote successful recombination, as DNA ends are highly recombinogenic in yeast (333).

3.3.2.2. Construction of pCAP03-derived capture vector (pSalbCAP)

The oligonucleotides designed as 50 bp capture arms for the target genomic region were cloned into pCAP03 (332) via Gibson assembly (334) (method described in section 6.4.1.1.). Two colonies were obtained on selective agar, from which plasmid DNA was isolated and screened by PCR for presence of the expected insert. In each case, a band of the expected size (364 bp) was observed (**Figure 28**), indicating that the capture arms

had been successfully incorporated into pCAP03 to yield pSalbCAP. Plasmid DNA was also sequenced to confirm that the capture arms were of the expected sequence (**Appendix Figure 87**).

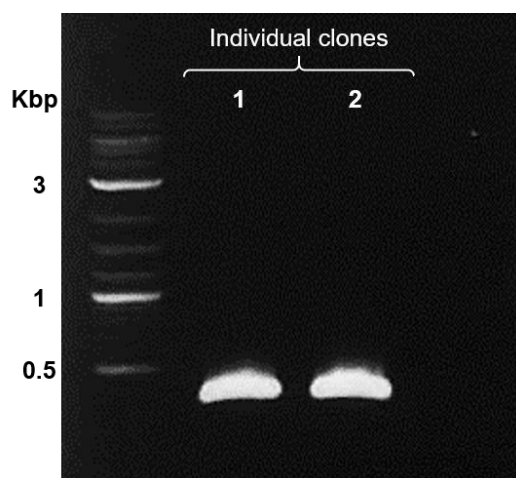


Figure 28: PCR screen of Image of plasmids isolated after Gibson assembly of pSalbCAP. Agarose gel image shows the amplified DNA fragment with the expected size (364 bp) to indicate successful incorporation of the capture arms into the pCAP03 backbone.

3.3.2.3. Spheroplast transformation and screening of cluster capture

Once the capture vector was constructed, the Moore 2.0 TAR cloning protocol (309) was followed to capture the target genomic region from digested *S. albus* J1074 gDNA via spheroplast transformation of *Saccharomyces cerevisiae* VL6-48N (methods described in sections 6.4.1.2. and 6.4.1.3.). In order to confirm successful capture of this genomic region to yield pCAPSalbC, a two-step PCR screening process was carried out, as a large number of colonies were obtained following negative selection with 5-FOA. First, five batches of 20 individual colonies were pooled into five single PCR reactions, for amplification of a target internal DNA fragment. Two of these pooled samples appeared to yield a positive result based on amplification of a DNA fragment of the expected size (~400 bp) (**Figure 29**). The 20 individual colonies from one of these pooled samples were then individually screened by PCR. This revealed that eight of these 20 colonies contained the successfully captured gene cluster based on amplification of a target internal fragment (**Figure 30**). To further confirm that the whole target genomic region had been captured, an analytical restriction digest was carried out on plasmid DNA purified from six individual clones, using restriction enzymes HindIII and SrfI that cut in various locations in the pCAPSalbC construct. This yielded the expected pattern of DNA bands following agarose gel electrophoresis, indicating that the full 18.5 kbp fragment had been successfully cloned (**Figure 31** and **Appendix Figure 86**).

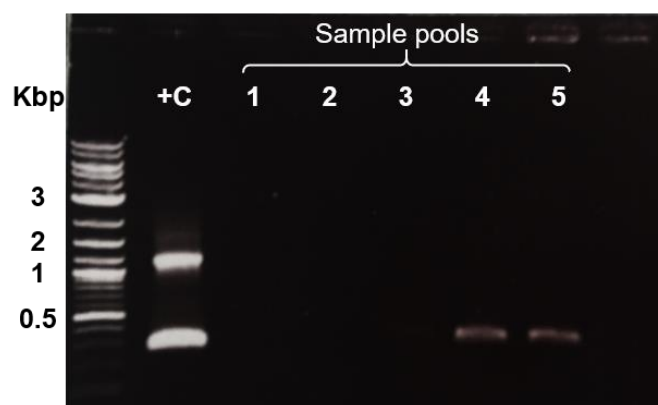


Figure 29: PCR screen of pooled colonies following TAR cloning. Agarose gel image shows amplification of DNA fragments from samples 4 and 5 which are the expected size of the target internal fragment (~400 bp), indicating successful capture of the *S. albus* J1074 gene cluster, yielding pCAPS**al**bC. +C= positive control (*S. albus* J1074 genomic DNA).

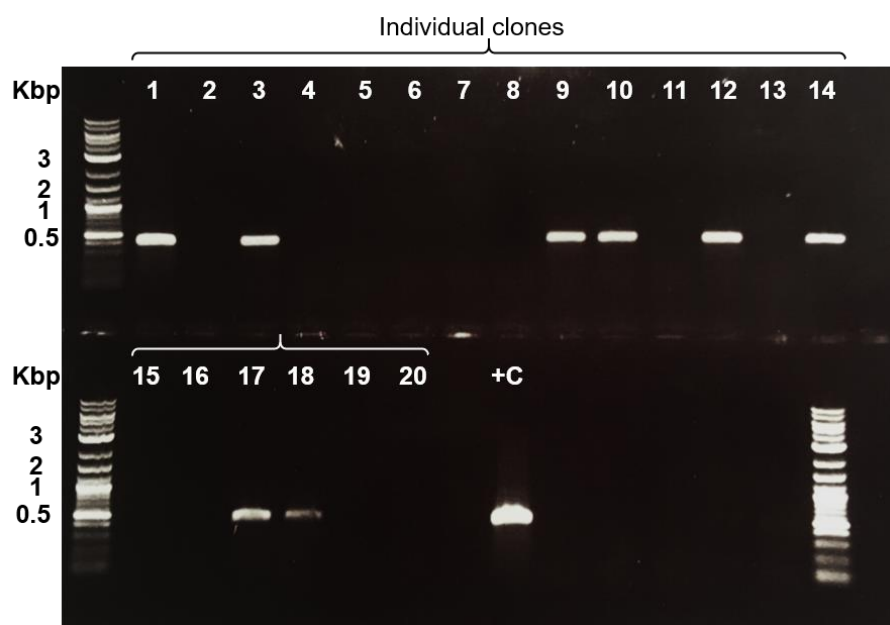


Figure 30: PCR screen of individual colonies following pooled batch screening. Agarose gel image shows amplification of DNA fragments obtained from colonies 1, 3, 9, 10, 12, 14, 17 and 18 which are the expected size of the target internal fragment (~400 bp), indicating successful capture of the *Streptomyces albus* J1074 biosynthetic gene cluster, yielding pCAPS**al**bC. +C= positive control (*S. albus* J1074 genomic DNA).

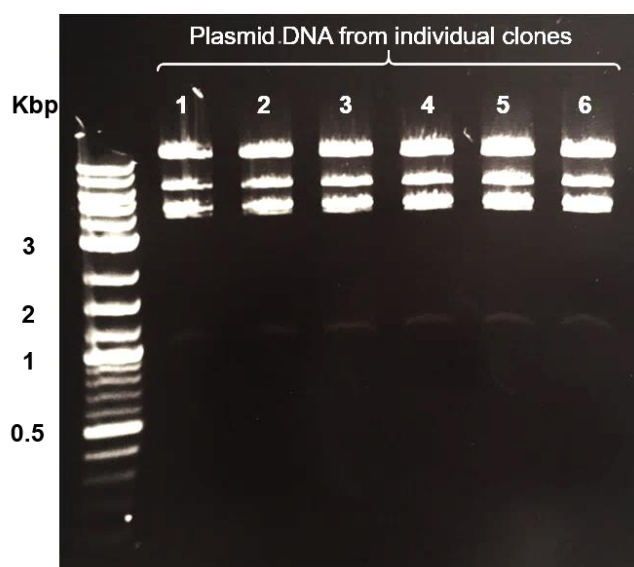


Figure 31: Analytical restriction digest of pCAPSalbC construct. Agarose gel image shows each plasmid sample digested with *SrfI* and *HindIII* yielding the five expected DNA fragments, indicating successful capture of the whole biosynthetic gene cluster from *Streptomyces albus* J1074 into pCAP03. The fragments at ~4 kbp appear as overlapped bands on the agarose gel.

3.3.3. Heterologous Expression and Metabolomic Screening

After confirming successful capture of the *S. albus* J1074 gene cluster into the pCAP03 vector, the next step was to introduce this construct into a heterologous host for expression and metabolomic analysis.

3.3.3.1. Deletion of precursor peptide gene from pCAPSalbC

In order to create a negative control for metabolomic screening, PCR targeting (309) was employed to engineer an individual in-frame deletion of the precursor peptide gene (*amiA*) within the pCAPSalbC construct (method described in section 6.4.4.). Individual colonies obtained after antibiotic selection were picked for plasmid extraction and the DNA was sequenced to confirm successful deletion of the precursor peptide gene. **Figure 32** shows a screenshot of the sequencing alignment showing that the precursor gene had been deleted and replaced with the expected 81 bp scar region, whilst the in-frame start and stop codons of the precursor peptide were retained.

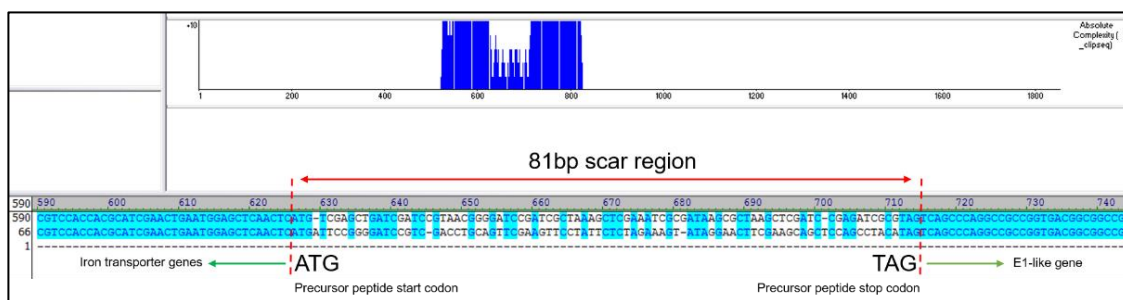


Figure 32: Sequence alignment showing deletion of precursor peptide gene from the pCAPSalbC. Top sequence corresponds to a fragment of the 'wild type' *S. albus* J1074 gene cluster where the precursor peptide gene has been replaced with a random 81 bp sequence (hypothetical 'scar' region) between the start and stop codons. Bottom sequence corresponds to the sequence obtained following Sanger sequencing of the pCAPSalbC Δ *amiA* mutant. Alignment of the two sequences indicate that the gene has been successfully deleted.

3.3.3.2. Heterologous expression

The construct containing the whole biosynthetic gene cluster from *S. albus* J1074 (pCAPSalbC) and the mutant construct lacking the precursor peptide gene (pCAPSalbC Δ *amiA*) were introduced via conjugation into four heterologous host strains: *Streptomyces lividans*, *Streptomyces laurentii*, *Streptomyces scabies* and *Streptomyces coelicolor* M1146 (method described in section 6.4.2.). I tested multiple heterologous hosts in order to increase the chances of observing pathway-related metabolite production. These particular strains were chosen as they have been previously utilised as heterologous hosts, and their genomes do not naturally contain the RiPP gene cluster of interest. Metabolomic screening was carried out with all four hosts, but the strain that

yielded an observable difference in metabolite production between the gene cluster and precursor peptide mutant was the *Streptomyces* superhost *S. coelicolor* M1146 (336). This strain was therefore used as a host for all further experiments.

3.3.3.3. Untargeted metabolomic screening

In order to identify pathway-related metabolites, wide-scale fermentations and liquid chromatography-mass spectrometry (LC-MS)-based screening were carried out (methods described in sections 6.2.4. and 6.4.7.). The two strains constructed (*S. coelicolor* M1146-pCAPSalbC and *S. coelicolor* M1146-pCAPSalbC Δ amiA) were grown in a range of culture media (BPM, R5, SM14) in order to identify optimal culture conditions for the pathway-associated metabolite. Associated metabolomic profiles were obtained following LC-MS analysis of culture extracts, taken at various time points over 10 days. These data were compared by overlaying MS base peak chromatograms and carrying out comparative metabolomic analyses using the Shimadzu Profiling Solution software. Initially, it was difficult to identify differences in metabolite production in any media tested, but upon screening in an additional medium (SM12), one particular metabolite could be observed from *S. coelicolor* M1146-pCAPSalbC but not the precursor peptide mutant, as observed after overlaying MS chromatograms from the two strains (**Figure 33**). This metabolite had a mass-to-charge ratio (m/z) of 272.16, which could be indicative of a small RiPP such as a tripeptide-derived metabolite.

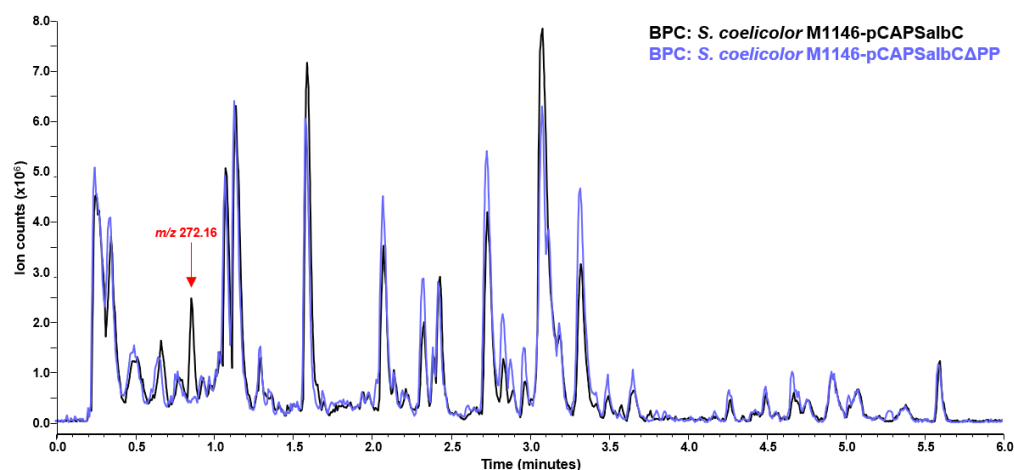


Figure 33: Comparison of base peak chromatograms (BPCs) from LC-MS analysis of *S. coelicolor* M1146-pCAPSalbC (black) and *S. coelicolor* M1146-pCAPSalbC Δ amiA (lilac). A peak with a retention time of 0.8 minutes corresponds to m/z 272.16, a metabolite produced by the strain expressing the whole cluster but not the precursor mutant. Data obtained on a Shimadzu IT-TOF.

Although this result looked interesting, a subsequent Profiling Solution analysis revealed that several other metabolites were also produced by the strain expressing the gene cluster (**Figure 34**). These metabolites eluted very early (less than 30 seconds into the LC-MS run) and were partially masked by other abundant metabolites such as sugars that elute within this early time period.

S. coelicolor M1146-pCAPSalbC replicas										S. coelicolor M1146-pCAPSalbCΔPP replicas					SM12 media control replicas	
Ion m/z	Ion RT	Ion Mass	PVal	1	2	3	4	5	1	2	3	4	5	1	2	
Compounds with early retention times	647.33	0.305	646.3189	1.50E-09	203260	186673	167596	228434	189120	0	0	0	0	0	0	
	532.25	0.315	531.238	1.30E-06	100256	159983	133946	89269	116256	0	0	0	0	0	0	
	526.29	0.323	525.2877	2.30E-09	104704	130302	140948	124682	143852	0	0	0	0	0	0	
	409.22	0.325	408.2094	1.80E-09	1032268	970795	992865	1204153	1252330	0	0	0	0	0	0	
	290.16	0.328	289.1509	3.90E-09	374867	367726	372913	394705	489671	0	0	0	0	0	0	
	510.27	0.333	509.2608	6.00E-08	1307321	1497729	1110801	958765	1162778	0	0	0	0	0	0	
	255.13	0.855	254.1235	7.00E-10	181075	221966	210496	186475	232244	0	0	0	0	0	0	
	m/z 272.16	0.858	271.1513	1.30E-12	2473857	2265867	2248234	2211799	2489214	0	0	0	0	0	0	
314.17	0.876	313.1607	2.70E-08	492061	527778	564969	560307	669575	0	0	0	0	0	130981		
268.16	1.651	267.1545	2.00E-06	222002	246320	217463	253794	375027	0	0	0	0	0	0		

Figure 34: Profiling Solution analysis of metabolites produced by *S. coelicolor* M1146-pCAPSalbC and *S. coelicolor* M1146-pCAPSalbCΔamiA. Data from five replica cultures of each strain are shown alongside two replicas of SM12 medium-only control cultures. Several metabolites with early retention times are produced by *S. coelicolor* M1146-pCAPSalbC but not *S. coelicolor* M1146-pCAPSalbCΔamiA, but these masses are masked by other major media components in the chromatograms shown in **Figure 33**. Data obtained on a Shimadzu IT-TOF.

In order to examine the fast-eluting metabolites in more detail, I repeated the LC-MS analysis using a polar C18 chromatography column, optimised for separation of polar metabolites, and applied a shallower methanol gradient to encourage later elution of the metabolites of interest (methods described in section 6.2.4.2.). This analysis confirmed that in addition to *m/z* 272.16, four other metabolites were produced by *S. coelicolor* M1146-pCAPSalbC but not the precursor peptide mutant. These metabolites were *m/z* 647.32 (also observed as [M+2H]²⁺, *m/z* 324.16), *m/z* 510.27, *m/z* 409.22 and *m/z* 338.18 (**Figure 35**), which could all represent larger pathway-related metabolites. Although this optimised chromatography approach led to the identification of putative pathway-related metabolites, further metabolite differences might still be missed based on limitations of the ionisation method and the solvent used for extraction. The newly-identified metabolites were examined by high-resolution LC-MS² analysis which indicated that they contain similar fragment ions, suggesting that these metabolites are structurally related (**Figure 36**). These metabolites might therefore represent a group of related metabolites all produced by the cloned biosynthetic gene cluster.

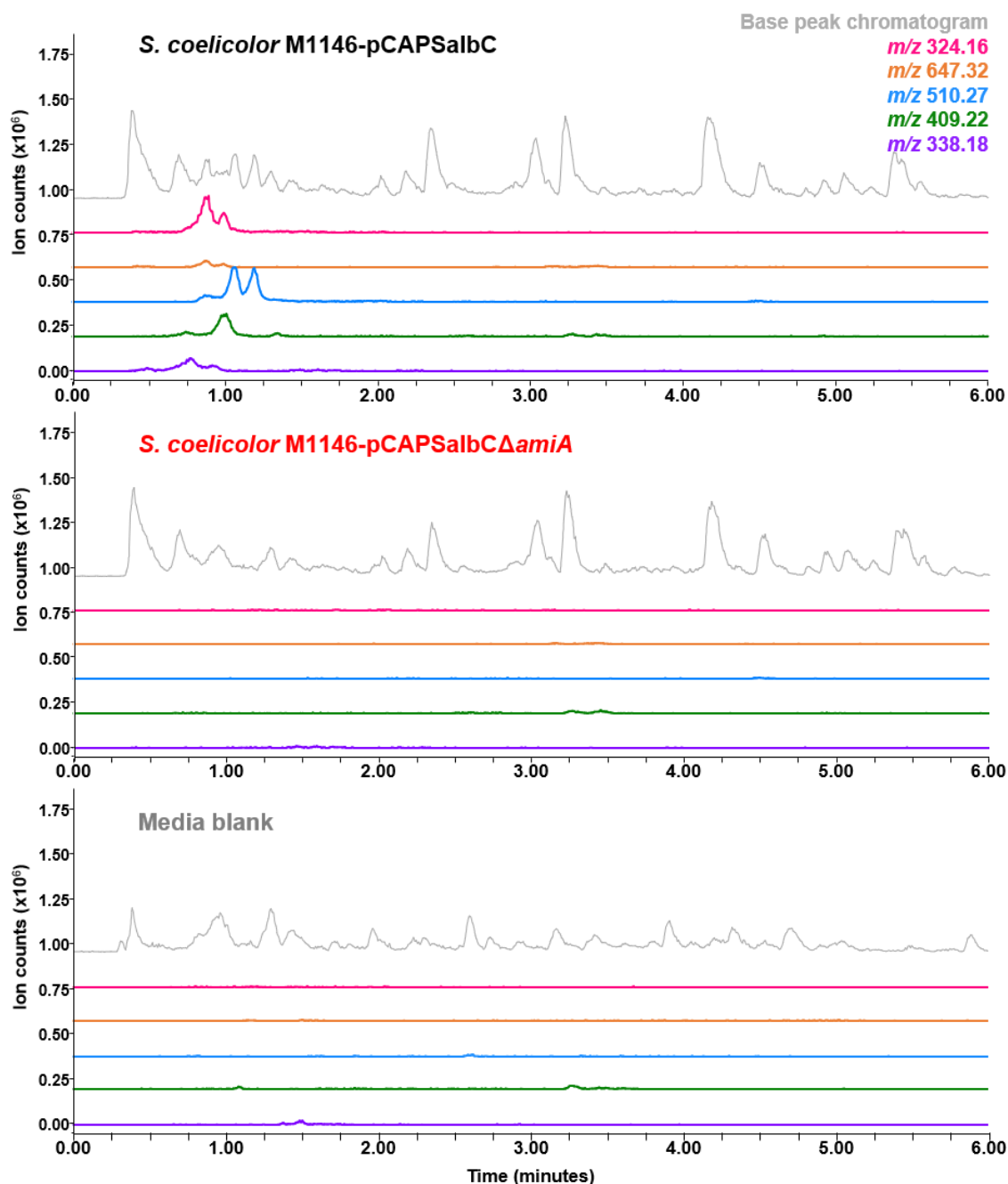


Figure 35: Comparison of extracted ion chromatograms of pathway-associated metabolites from *S. coelicolor* M1146-pCAPSalbC and *S. coelicolor* M1146-pCAPSalbC Δ amiA. The four identified cluster-associated metabolites are shown: m/z 647.32 (orange) (also observed as $[M+2H]^{2+}$ m/z 324.16 (pink)), m/z 510.27 (blue), m/z 409.22 (green) and m/z 338.18 (orange) with base shift applied. Base peak chromatogram is shown in grey. Data obtained on a Shimadzu IT-TOF.

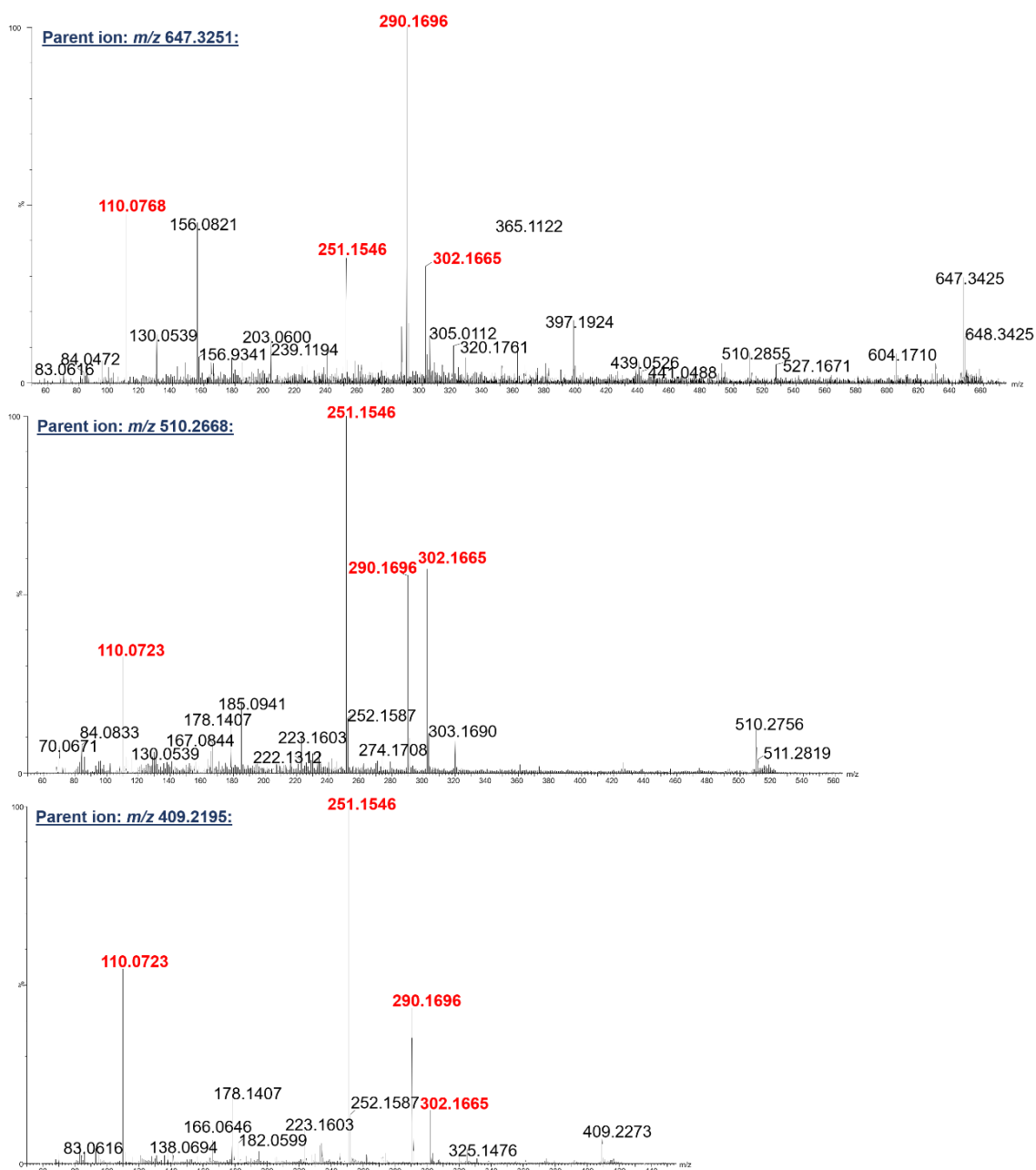


Figure 36: High-resolution LC-MS² data for pathway-related metabolites; m/z 647.32, m/z 510.27 and m/z 409.22, obtained using a Waters Synapt G2Si. Common fragment ions are highlighted in red.

Overall, these LC-MS analyses suggest that uncharacterised natural products that are polar in nature might be overlooked when standard metabolomic screening conditions are used that are not optimised for separation and detection of polar metabolites. It is also interesting that these metabolites were only observed in one of the media tested (SM12). This is a complex medium and there might be certain components that helped trigger production of the biosynthetic gene cluster. One of the reasons I chose to test SM12 is that it contains several protein-rich ingredients that I hypothesised would provide an abundant supply of amino acid building blocks to produce the target RiPP natural product.

3.3.4. Insertional Disruption of the *S. albus* J1074 Biosynthetic Gene Cluster

After identifying putative pathway products from heterologous expression of the model biosynthetic gene cluster, I was interested to see whether these same metabolites could be detected from the native producer- “wild type” *S. albus* J1074 (337)- using the same culture and metabolomic screening conditions. To obtain a negative control for metabolomic screening of *S. albus* J1074, I constructed a pathway mutant of the strain via insertional disruption of the biosynthetic gene cluster *in vivo* (methods described in section 6.4.3.). I chose this approach as although not as clean as obtaining a precise gene or whole cluster deletion, it was a quicker way to confirm the results already obtained from heterologous expression of the gene cluster.

3.3.4.1. Construction of disruption vector

To construct the *S. albus* J1074 mutant, I first cloned a DNA fragment corresponding to the genes encoding the hydrolase (*amiC*) and YcaO-domain protein (*amiD*) that physically overlap on the chromosome, ensuring that the predicted catalytic domains of the encoded proteins would be included in the disruption. Therefore, the gene cluster would be highly unlikely to be functional after recombining with this fragment. Once the disruption fragment was cloned, I ligated it into the suicide plasmid pKC1132 to create the construct pKCΔ*amiCD*. After transforming this construct into *E. coli* DH5α, individual colonies obtained on selective agar were screened by PCR to confirm presence of the insert which was ~2 kbp in size (**Figure 37**). Plasmid DNA was isolated from individual clones that were positive for the insert and sequenced to confirm the presence of the disruption fragment.

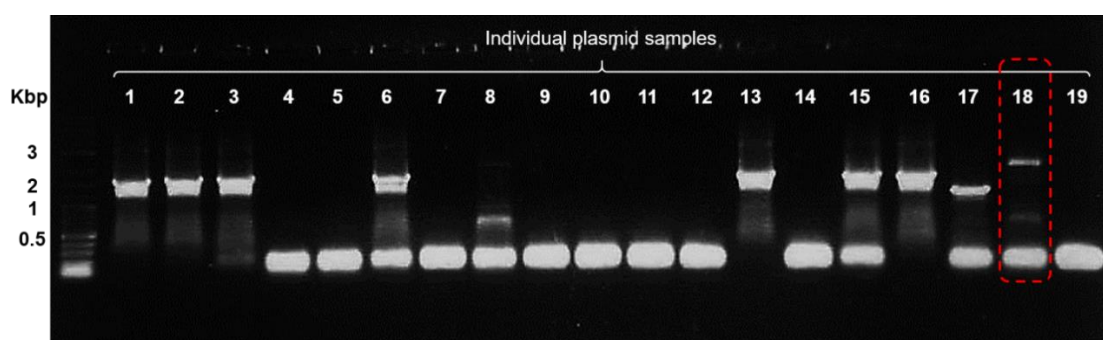


Figure 37: PCR screen of pKCΔ*amiCD* disruption construct. Agarose gel image shows amplification of fragments from individual *E. coli* colonies transformed with the ligated pKCΔ*amiCD* disruption construct. Lane 18 contains a band of ~2 kbp, which is the expected size of the target insert.

3.3.4.2. Conjugation of *S. albus* J1074 with pKCΔ*amiCD*

After confirming successful construction of pKCΔ*amiCD*, the vector was introduced into *E. coli* ET1256 via triparental mating with *E. coli* TOP10-pUZ8002 (described in section 6.2.1.3.) followed by intergenic conjugation with *S. albus* J1074 (described in section 6.2.2.3.). Successful single crossover between pKCΔ*amiCD* and the *S. albus* J1074 chromosome was confirmed by colony PCR using primers that amplify the intact *amiCD* fragment from wild type sequences but not from the pathway disrupted mutants. PCR screening showed that the intact fragment was amplified from *S. albus* J1074 gDNA (2 kbp), but not from pathway mutant exconjugants (*S. albus*-pKCΔ*amiCD*), although non-specific amplification was observed from these colonies (0.5 and 3 kbp) (**Figure 38**). This indicated successful disruption of the *S. albus* J1074 pathway.



Figure 38: PCR screen of *S. albus* J1074-pKCΔ*amiCD* exconjugants. Agarose gel image shows amplification of DNA fragments from individual colonies. The positive control in lane 1 (+C) shows amplification of the undisrupted gene region from wild type *S. albus* J1074. Samples 1-17 represent individual exconjugants from which amplification of a larger band indicate disruption of the gene region following insertion of pKCΔ*amiCD*.

3.3.4.3. Metabolomic screening of *S. albus* J1074 and *S. albus*-pKCΔ*amiCD*

Once the *S. albus*-pKCΔ*amiCD* mutant had been constructed, the strain was grown alongside wild type *S. albus* J1074 under the same conditions as for heterologous expression experiments, using the same experimental conditions for LC-MS analysis. The metabolomic profiles of *S. albus* J1074 and *S. albus*-pKCΔ*amiCD* were compared and interestingly, the putative pathway products identified from heterologous expression of the cluster were also detected from *S. albus* J1074 but not the *S. albus*-pKCΔ*amiCD* mutant. This included the four putative pathway-related metabolites *m/z* 647.32, *m/z* 510.27, *m/z* 409.22 and *m/z* 338.18. Interestingly however, the *m/z* 272.16 metabolite was detected in the *S. albus* mutant, and not the wild type strain (**Figure 39**). This could suggest that *m/z* 272.16 is not a main pathway product but might instead correspond to a biosynthetic intermediate produced by the mutant. The other four metabolites identified (*m/z* 647.32, *m/z* 510.27, *m/z* 409.22 and *m/z* 338.18) could be genuine products of the

cloned biosynthetic pathway, as they were detected from both the native and heterologous host of the biosynthetic gene cluster.

It is interesting that the newly identified metabolites were detectable from the native producer without any manipulation of native pathway elements. Cryptic biosynthetic gene clusters that have not previously been characterised are often only “switched on” in the native producer after the manipulation of pathway regulators or promoters. In the case of *S. albus* J1074, I was able to detect the putative pathway products after replicating favourable growth and screening conditions identified during the previous heterologous screening experiments.

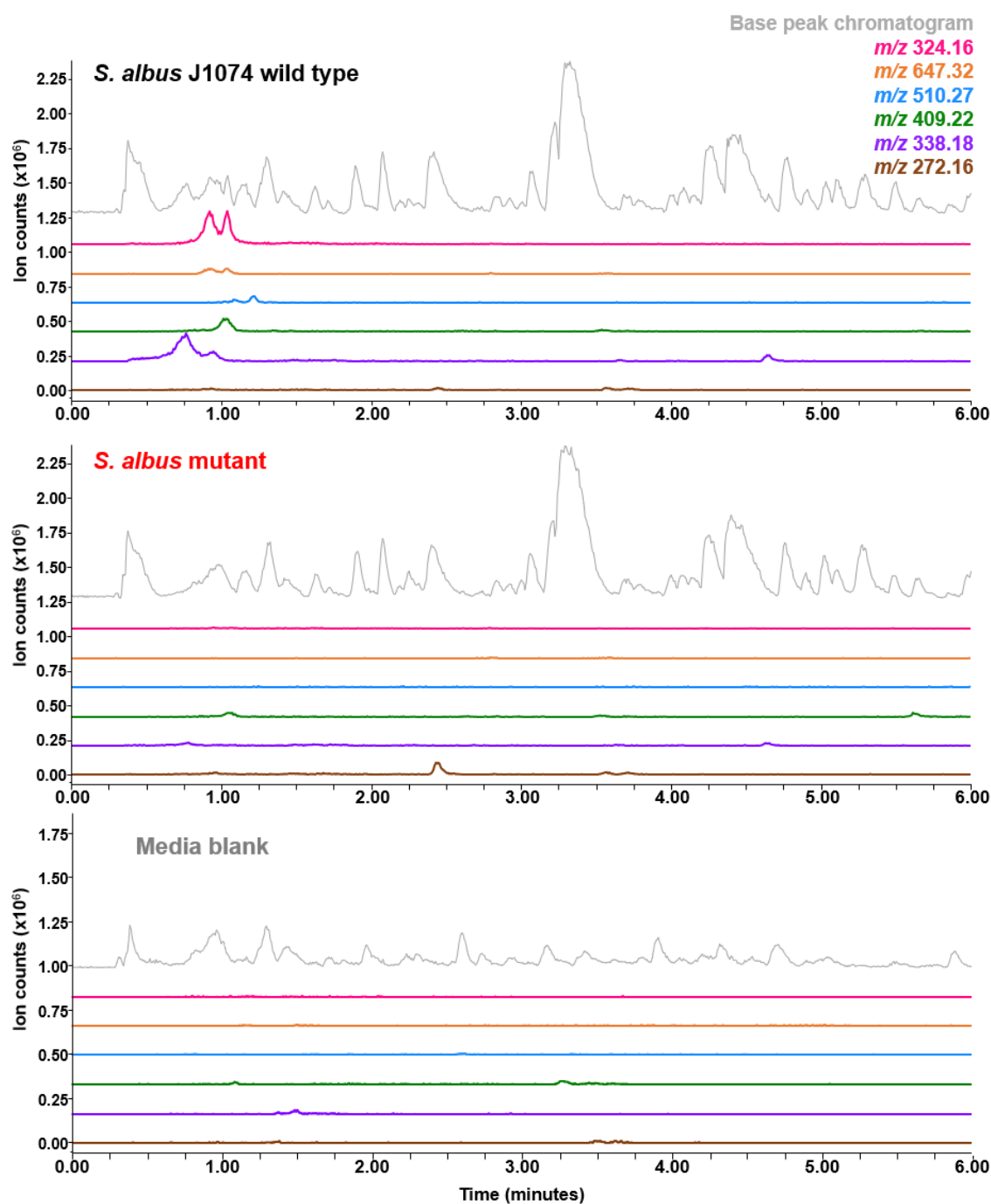


Figure 39: Comparison of extracted ion chromatograms of pathway-associated metabolites from *S. albus* J1074 wild type and *S. albus*-pKCΔamiCD. Five cluster-associated metabolites are shown: *m/z* 647.32 (orange) (also observed as $[M+2H]^{2+}$ *m/z* 324.16 (pink)), *m/z* 510.27 (blue), *m/z* 409.22 (green), *m/z* 338.18 (orange) and *m/z* 272.16 (brown) with base shift applied. Base peak chromatogram is shown in grey. Data obtained on a Shimadzu IT-TOF.

3.3.5. Comparing Metabolites Produced by the Native and Heterologous Hosts

In order to confirm that the metabolites detected from the native and heterologous hosts were indeed the same metabolites, I compared the MS² data of the metabolite with the largest mass (observed as m/z 324.16 [M+2H]²⁺) from both *S. coelicolor* M1146-pCAPSalbC and *S. albus* J1074. Although the LC-MS total ion current (TIC) peak shapes differ slightly in appearance, the identical fragmentation patterns observed in both samples suggest that the identified metabolite is structurally identical (**Figure 40**).

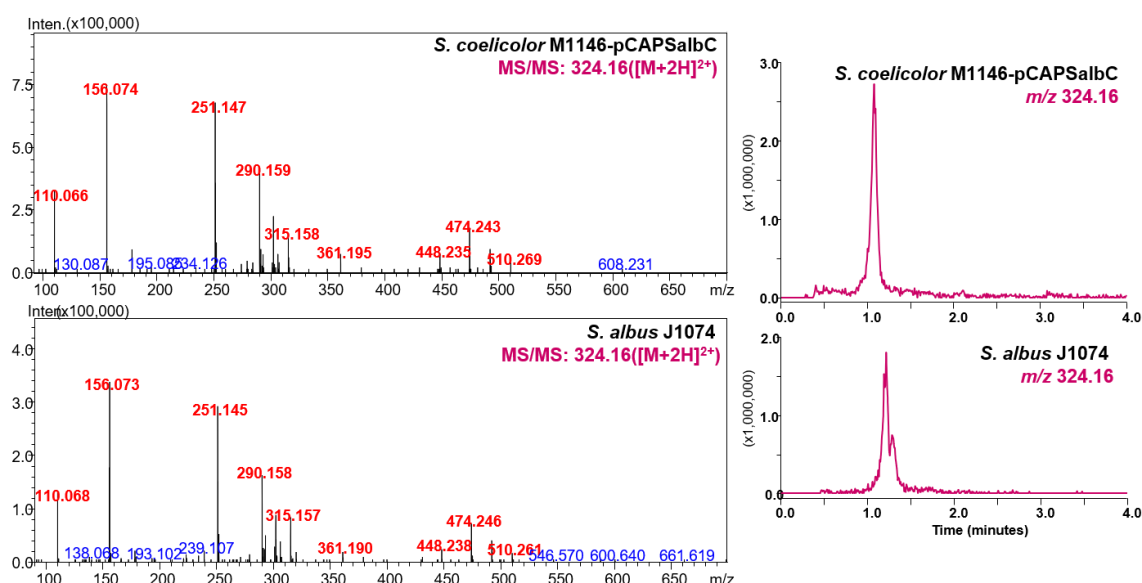


Figure 40: Comparison of MS² data for m/z 324.16 [M+2H]²⁺ from *S. coelicolor* M1146-pCAPSalbC *S. albus* J1074). Data obtained using a Shimadzu IT-TOF.

3.3.5.1. GNPS networking analysis

The putative pathway-related metabolites were identified from screening trials with both the native producer and heterologous host of the target biosynthetic gene cluster. LC-MS data files relating to both of these strains were submitted to GNPS for metabolomic networking (described in section 2.1.2.2.). One of the resulting molecular networks connected many of the pathway-related metabolites identified during fermentation trials (m/z 510.27, m/z 409.22 and m/z 338.18). This suggested that these metabolites are structurally related based on the presence of similar fragmentation patterns (**Figure 41**). Although the larger identified metabolite (m/z 647.32) was not included in this network, the presence of similar fragmentation patterns observed from high-resolution LC-MS² analysis between m/z 647.32, m/z 510.27 and m/z 409.22 (**Figure 36**) suggest that the four metabolites identified are structurally related.

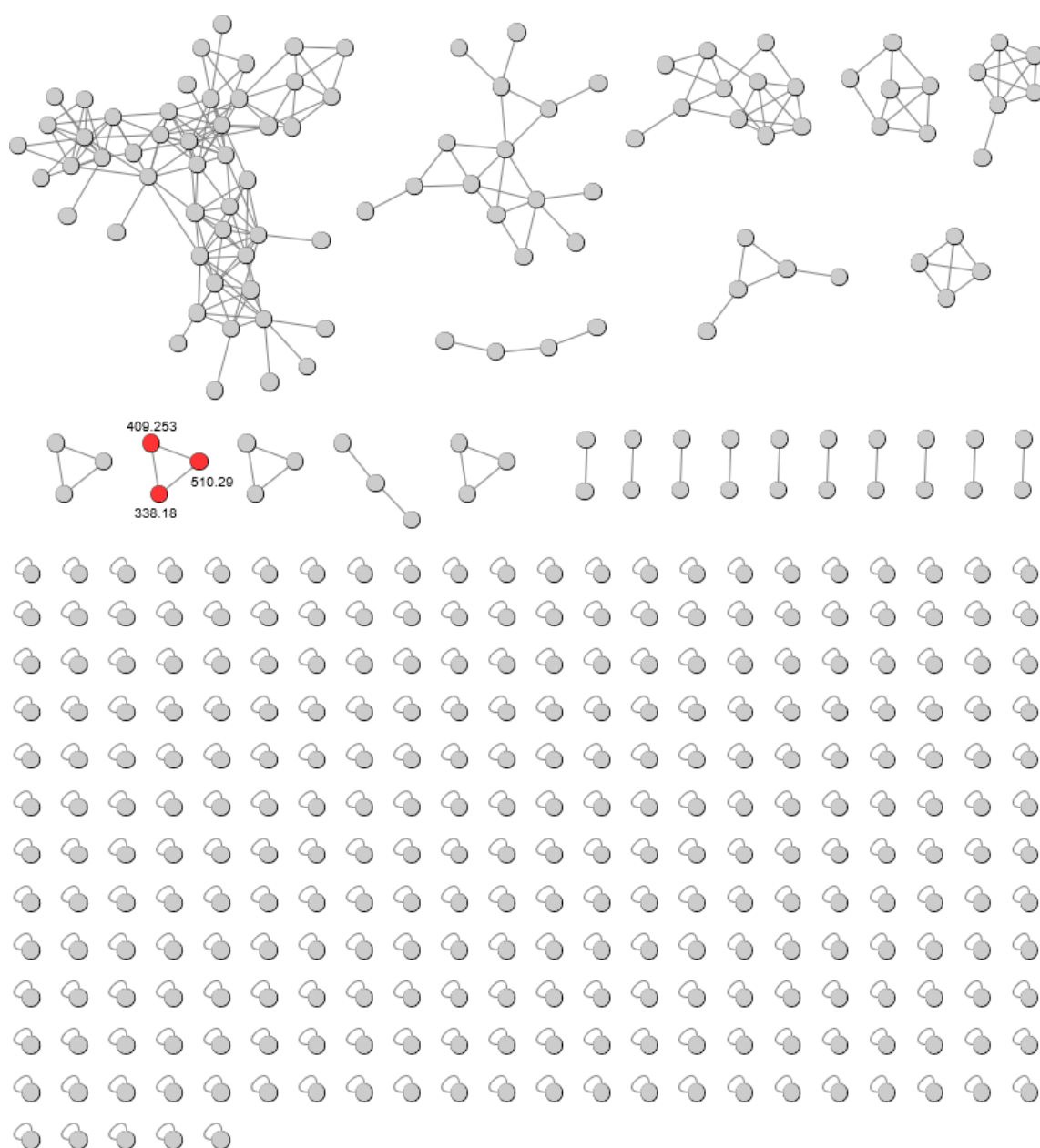


Figure 41: Global Natural Products Social (GNPS) molecular network of pathway-associated metabolites. Obtained from analysis of a range of LC-MS data files obtained from fermentation of *S. coelicolor* M1146-pCAPSalbC and *S. albus* J1074. The sub-network corresponding to the identified pathway-related metabolites is highlighted in red.

3.3.6. Gene Deletions from the TAR Cloned *S. albus* J1074 Gene Cluster

3.3.6.1. Deletion of biosynthetic genes from pCAPSalbC

In order to further confirm that the newly identified metabolites were genuine products of the cloned gene cluster, and to investigate the involvement of individual biosynthetic genes, I constructed further gene deletions from the pCAPSalbC construct. Theoretically, any gene captured into the TAR cloned fragment could be involved in biosynthesis of the identified metabolites, therefore I deleted all genes that form the proposed biosynthetic gene cluster as well as the additional captured genes. As with the precursor peptide deletion, this was achieved using PCR targeting (309) (method described in section 6.4.4.). Individual deletions were made of genes encoding the oxygenase, MarR regulator (*amiM*), oxidoreductase (*amiX*), E1-ubiquitin-like enzyme (*amiB*), hydrolase (*amiC*), YcaO-domain protein (*amiD*), dehydrogenase (*amiE*) and peptide methionine reductase proteins. The four iron transporters were deleted as a combined genetic fragment, as were the two ABC transporters. This is because these genes are transcriptionally coupled, and each comprise a single transport system. The two transcriptionally coupled acetyltransferase genes were also deleted as a combined fragment. After PCR screening and DNA sequencing to confirm deletion of these genes from pCAPSalbC, the ten resulting deletion constructs were introduced into *S. coelicolor* M1146 via conjugation and the resulting mutant strains were fermented and screened under the same conditions as for the previous heterologous expression experiments.

3.3.6.2. Metabolomic screening of gene deletion mutants

LC-MS analysis of each mutant strain indicated that several gene deletions abolished production of the previously identified metabolites, suggesting that these genes encode essential biosynthetic proteins. These include the iron transporters (AmiF1-4), E1-ubiquitin-like enzyme (AmiB), hydrolase (AmiC), YcaO-domain protein (AmiD) and the dehydrogenase (AmiE) (**Figure 42** and **Figure 43**). I therefore proposed that these genes, along with the precursor peptide gene (*amiA*), make up a minimal biosynthetic gene cluster for the identified metabolites (**Figure 44**). The oxidoreductase gene deletion had an effect on streptamidine production but did not completely abolish biosynthesis (**Figure 42**). Gene deletions that had no effect on production of the identified metabolites included those encoding the oxygenase, MarR regulator, ABC transporters, peptide methionine reductase and acetyltransferases. Interestingly, some of the mutants appeared to over-produce some additional metabolites as observed when comparing their metabolomic profiles with that of *S. coelicolor* M1146-pCAPSalbC (carrying the intact gene cluster). These could be indicative of potential biosynthetic intermediates or shunt metabolites. One of these metabolites was *m/z* 272.16, which was previously

observed from the pathway-disrupted *S. albus* mutant. The other shunt metabolites observed include m/z 314.17, m/z 332.18 and m/z 354.16. The mass differences between these metabolites suggest they might be structurally related: a difference of 18.01 between 314.17 and 332.18 reflects hydration, and the difference of 21.98 between 332.18 and 354.16 indicates formation of a sodium adduct instead of a proton adduct. The importance of these shunt metabolites in biosynthesis of the fully modified RiPP product is discussed further in Chapter 4.

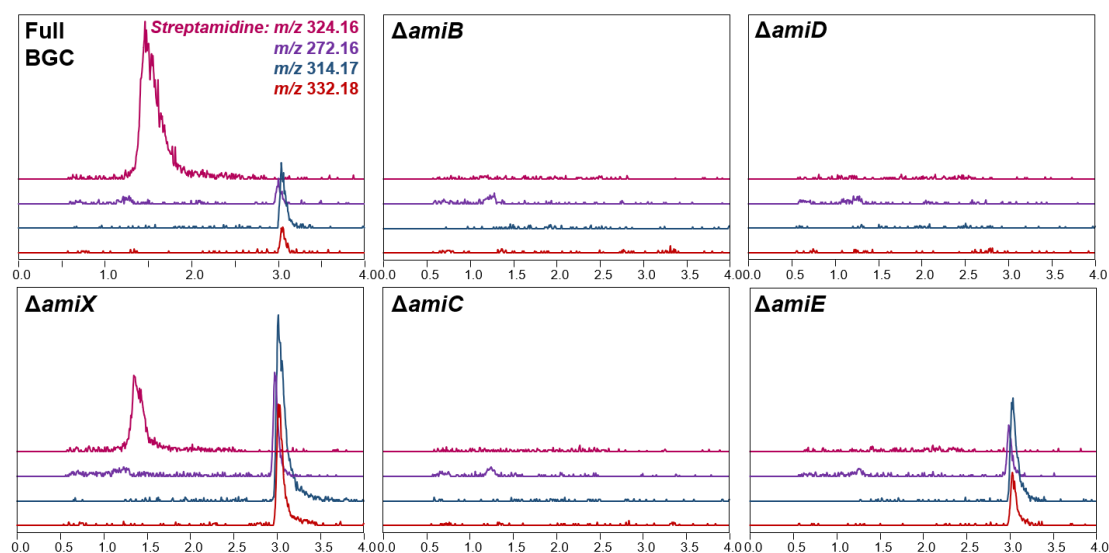


Figure 42: LC-MS analysis of key biosynthetic gene deletions from pCAPSalbC. Extracted ion chromatograms indicate production of a putative key product of the *S. albus* J1074 biosynthetic gene cluster (m/z 324.16) compared with other over-produced metabolites that might be indicative of pathway intermediates (m/z 272.16, m/z 314.17 and m/z 332.18). Data obtained on a Shimadzu IT-TOF.

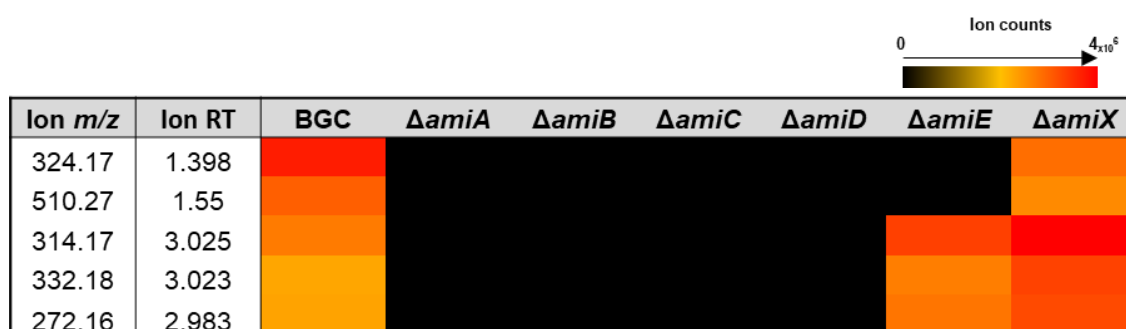


Figure 43: Heat map of metabolite production in pathway mutants. Colour scale indicates the level of production of pathway-associated metabolites from *S. coelicolor* M1146-pCAPSalbC compared with individual pathway mutants. Data obtained on a Shimadzu IT-TOF and analysed by Profiling Solution. The ion intensities of metabolites produced by each mutant strain in triplicate were combined, averaged and converted into a colour scale. Gene designations are as shown in figure 44.

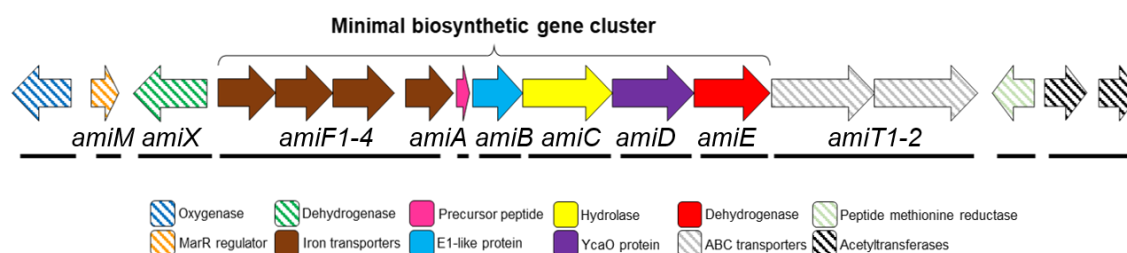


Figure 44: Genetic region TAR cloned from *S. albus* J1074. The minimal biosynthetic gene cluster responsible for production of pathway-related metabolites is indicated, as identified from metabolomic screening of gene deletion mutants. Essential biosynthetic genes are indicated with filled arrows and nonessential biosynthetic genes are indicated with striped arrows.

3.3.7. Genetic Complementation of Gene Deletions

3.3.7.1. Construction of gene complementation vectors

In order to confirm that the results from gene deletions were genuine and not due to off-target effects, the deletion mutants that abolished metabolite production were genetically complemented (methods described in section 6.4.5.). The genes for the precursor peptide (*amiA*), E1-like protein (*amiB*), hydrolase (*amiC*), YcaO-domain protein (*amiD*) and dehydrogenase (*amiE*) were cloned into the expression vector pIJ10257, which contains an *ermE** promoter. After PCR screening and DNA sequencing to confirm presence of the correct inserts, each construct was conjugated into the corresponding deletion mutant of *S. coelicolor* M1146-pCAPSalbC. Complementation was successful with all of these strains except for the precursor peptide mutant. This might be because precursor peptides in RiPP gene clusters have very specific transcriptional and translational regulation, and therefore making deletions in this region might disrupt this important functionality. To overcome this, I attempted to construct alternative complementation constructs for the precursor peptide, cloning longer stretches of sequence upstream of the precursor peptide into pIJ10257, in case an important promoter region was previously omitted. However, these constructs were not able to complement the gene deletion. Difficulty with complementing RiPP precursor peptide deletions has also been observed by other members of the lab group with the bottromycin and thiovarsolin pathways (124,331).

3.3.7.2. Metabolomic screening of genetically complemented mutants

Successful complementation of the YcaO, E1-like and dehydrogenase deletions resulted in restored production of the previously identified pathway metabolites (**Figure 45**). This suggests that the abolished production previously observed was genuinely caused by the gene deletions and not off-target effects in the mutant. Complementation of the hydrolase gene deletion however did not restore production of the putative pathway metabolites. I also considered whether there might be a problem with the original

hydrolase deletion construct design. As the hydrolase gene lies between two other co-transcribed biosynthetic genes (**Figure 44**), deletion of *amiC* could have disrupted expression of these neighbouring pathway genes. I therefore re-designed the hydrolase gene deletion construct, reducing the size of the deleted region to ensure that promoter regions of nearby genes were not interrupted. However, this did not change the metabolomic profile of the hydrolase mutant or the result of the subsequent complementation.

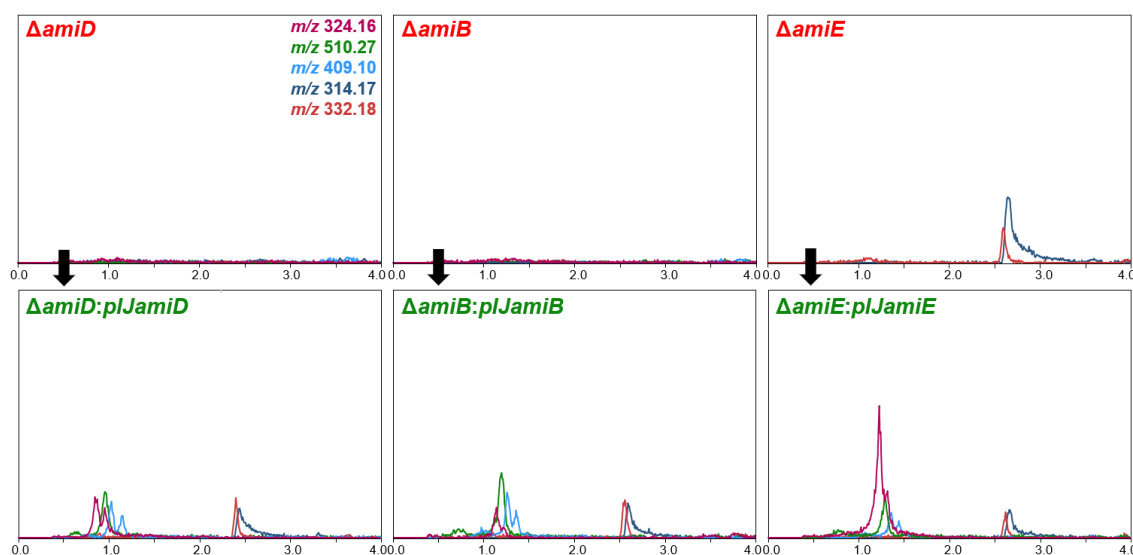


Figure 45: LC-MS analysis of genetic complementations of deletions from the pCAPSalbC construct. Extracted ion chromatograms of pathway-associated metabolites are indicated, (*m/z* 324.16 (pink), *m/z* 510.27 (blue) and *m/z* 409.22 (green)) as well as extracted ion chromatograms of putative biosynthetic intermediates (*m/z* 314.17 (dark blue) and *m/z* 332.18 (red)). Data obtained on a Shimadzu IT-TOF.

3.4. Chapter Summary

This chapter presents a range of genetic approaches that were undertaken to clone and study a model RiPP biosynthetic gene cluster from *S. albus* J1074. The genomic region containing the cluster was TAR cloned and heterologously expressed to identify putative pathway products, which were also detected from the native producer. A series of gene deletions were constructed to confirm which biosynthetic enzymes were involved in producing the identified metabolites, some of which were genetically complemented to confirm that these proteins were essential for biosynthesis. Together these results suggest that the four metabolites identified from LC-MS analyses (*m/z* 647.32, *m/z* 510.27, *m/z* 409.22 and *m/z* 338.18) were genuine products of the previously uncharacterised RiPP biosynthetic gene cluster from *S. albus* J1074, and I could propose a minimal set of genes required for biosynthesis.

Overall, the results presented in this chapter demonstrate that previously unknown biosynthetic gene clusters identified by genome mining can be subsequently characterised through cloning, genetic manipulation and metabolomic screening of the pathway. In the future these approaches can continue to be applied to other newly identified gene clusters to help characterise novel metabolites. I also showed that by trialling different growth media, metabolomic screening conditions and chromatography methods, it is possible to identify metabolites that were previously not detected. This is particularly interesting in the case of *S. albus* J1074, which is an extensively studied strain that is commonly used as a heterologous host (299,300,329,338–341), and yet the metabolites identified in this study have not been previously reported. Variations in growth and screening conditions should therefore also be considered when screening other novel biosynthetic gene clusters in the future.

The experiments described in this chapter led to the identification of a group of metabolites that appear to be genuine products of the novel biosynthetic gene cluster from *S. albus* J1074. The following chapter will describe the characterisation of these metabolites in further detail.

Chapter 4: Investigating the Structure, Biosynthesis and Activity of a Novel RiPP

4.1. Introduction

The data presented in Chapter 3 describe the identification of metabolites produced by a novel RiPP biosynthetic gene cluster. This chapter will focus on characterisation of the putative final product of this gene cluster, using a combination of purification techniques, structural analyses and biological investigations. Structural elucidation of natural products is an important aspect of their characterisation as it can provide clues about the biosynthetic steps of the associated natural product pathway. Another key goal of natural products research is to identify the biological activities of novel metabolites, which is also tightly linked to the molecular structure.

4.1.1. Structural Characterisation of Natural Products

Many different spectroscopic methods have been developed to aid structural elucidation work in organic chemistry. Early work involved X-ray crystallography, pioneered by Max von Laue in 1912, who discovered that bombarding crystals with X-rays resulted in characteristic diffraction patterns on a photographic plate (342). The use of this technique to solve molecular structures was widely demonstrated by Dorothy Crowfoot Hodgkin during the 1940s-'60s, who solved the structure of important metabolites such as penicillin (343), vitamin B₁₂ (344), insulin (345) and thiostrepton (140). More recently, nuclear magnetic resonance (NMR) has been widely adopted to solve structures of small metabolites. NMR provides information about the chemical shifts of atoms as atomic nuclei exhibit characteristic nuclear spin when exposed to electromagnetic radiation. The measurement of nuclear magnetic moments using magnetic resonance absorption was first demonstrated in 1938 by Rabi (346). 2D NMR was introduced in 1976 by Ernst (347), which provides further correlation data between atoms in a structure. A more recent technique developed for structural elucidation of small metabolites is microcrystal electron diffraction (MicroED), which utilises CryoEM (electron cryo-microscopy) to solve structures of small metabolites (348). Solid ground powder is first applied to glass cover slides, which are then deposited on a carbon-copper grid, flash-frozen in liquid nitrogen and transferred to a cryoelectron microscope. Nanocrystals on the grid surface, approximately 100 nm in diameter, are continuously rotated at 140 degrees and imaged

to create a movie. From this, structural information is inferred. Structures can be solved to 1 Å resolution with data from a single nanocrystal (349).

4.1.2. Biological Characterisation of Natural Products

In terms of investigating antimicrobial activities of metabolites, a commonly used method to investigate antimicrobial activity is the agar disk-diffusion method, first developed in 1940 to study the activity of penicillin (350). This method can be used to test the inhibitory activity of metabolites or extracts against a range of pathogens or indicator strains. This can give an initial indication of antimicrobial activity, but more accurate minimum inhibitory concentration (MIC) testing can be achieved using dilution methods (83), where the MIC is determined as the lowest concentration of antimicrobial agent that completely inhibits growth of the test organism (351). To overcome issues with ambiguity or subjectivity, several colorimetric-based assays have also been developed, such as the Alamar blue dye (resazurin) assay (352). In order to investigate other biological activities such as metal-binding, other assays have also been developed. A universal method for determining siderophore activity is the chrome azurol S (CAS) assay. Here, CAS and hexadecyltrimethylammonium bromide (HDTMA) form a tight complex with iron that produces a blue colour. When a siderophore is added, iron is removed from the complex resulting in a colour change from blue to orange (353). This colorimetric assay can be carried out on agar plates or in liquid, and has been further developed to investigate the chelation of other metal ions (354). Other assays have also been developed to detect the chelation of other metals such as the 4-(2-pyridylazo)-resorcinol (PAR) assay that detects zinc binding (355).

4.2. Chapter Aims

The pathway-related metabolites identified from metabolomic screening experiments (Chapter 3) are hypothesised to represent novel RiPP metabolites. As the biosynthetic gene cluster was previously uncharacterised, these RiPPs could harbour unique structural features or novel bioactivity. The overall aim of this chapter was to characterise these newly identified metabolites. To achieve this, the specific objectives of this chapter were:

- (i) Utilise metabolomic prediction tools to identify a putative core peptide from the metabolomic data
- (ii) Purify the putative final metabolite of the biosynthetic pathway
- (iii) Elucidate the structure of the purified metabolite using NMR
- (iv) Investigate the biosynthesis of the identified metabolite
- (v) Investigate the biological activity of the metabolite

4.3. Results and Discussion

4.3.1. Core Peptide Prediction and Structural Hypotheses

As discussed in previous chapters, analysis of MS and fragmentation data can help provide clues about the structural identity and amino acids present in a peptide. Before purifying the final pathway-related metabolite for structural elucidation, I made some initial structural and core peptide predictions based on the MS data I had already obtained.

4.3.1.1. Core peptide prediction

Considering the pathway-related metabolites that have been identified, I was interested to see how these might relate to a modified core peptide within the *S. albus* precursor peptide AmiA. To achieve this, I utilised a RiPP peptide mass calculator developed at the John Innes Centre by Andrew Truman and Govind Chandra (unpublished). I submitted the *S. albus* J1074 precursor peptide sequence along with the identified singly charged metabolite masses (m/z 647.32, m/z 510.27, m/z 409.22 and m/z 338.18). I then predicted the possible mass changes that might occur based on the biosynthetic enzymes in the pathway. This included mass losses such as 18.02 (corresponding to a dehydration that could be catalysed by the oxidoreductase amiX) and 2.02 (corresponding to dehydrogenation that could be catalysed by the dehydrogenase AmiE). Interestingly, after submitting these data for analysis, the calculator mapped all the identified metabolites to within the same six-residue region of the precursor peptide: His-Leu-Ser-Ala-Thr-His. The metabolites related to this sequence as follows: m/z 647.32 (HLSATH), m/z 510.27 (HLSAT), m/z 409.22 (HLSA) and m/z 338.18 (HLS) (**Appendix Figure 88**). This suggested that these masses relate to the same RiPP metabolite, with the smaller masses representing breakdown products or intermediates of the final metabolite. Going forward, I proposed that the final product of the pathway is a metabolite with m/z 647.32 (also observed as m/z 324.16 $[M+2H]^{2+}$), as this was the largest mass identified. The peptide calculator indicated that several possible combinations of mass changes could theoretically match the final modified metabolite with the proposed core sequence, therefore the exact post-translational modifications could not be determined from this prediction tool.

4.3.1.2. Initial structural hypothesis

One of the possible mass changes calculated was for a single dehydration to occur within the proposed core sequence. Based on the biosynthetic genes present in the gene cluster, it would be plausible for a single dehydration to occur with the formation of an

oxazoline heterocycle, which would be consistent with ATP-dependent cyclodehydration catalysed by the YcaO-domain protein with the E1-like partner protein. This could take place on either the threonine or serine residue (**Figure 46**). However, the presence of additional dehydrogenases in the cluster would therefore seem redundant. The gene deletion experiments described in Chapter 3 indicated that one of the dehydrogenase enzymes appeared to be essential for biosynthesis of the final metabolite. It could therefore be possible that the oxazoline is oxidised to its corresponding oxazole, but this would require a further reduction in the metabolite to account for these mass changes. This could occur on one of the histidine rings for example, but this would be an unusual post-translational modification for a RiPP (**Figure 46**).

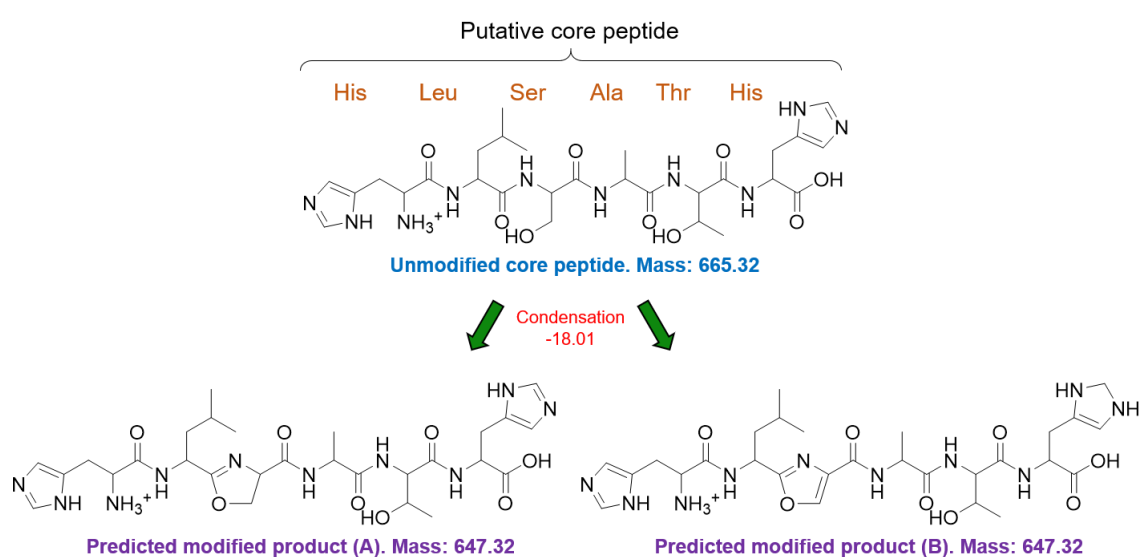


Figure 46: Structural hypotheses for the putative final pathway metabolite (m/z 647.32). Structures based on exact mass, core peptide predictions and post-translational modification predictions.

4.3.1.3. Molecular formula prediction

The predicted modified peptides shown in **Figure 46** have a chemical formula of $C_{28}H_{42}N_{10}O_8$. According to the Shimadzu formula predictor tool, this is the highest scoring and therefore most likely chemical formula based on the ion peaks of the putative metabolite (m/z 647.32). This formula is also supported by accurate mass data for this metabolite obtained on a Synapt G2Si (calculated $[M+H]^+$ m/z 647.3260; observed m/z 647.3251, **Appendix Table 26**). This could support the structural predictions presented, but these atoms could also be arranged in different ways. The exact structure of this RiPP therefore needed to be confirmed following purification of the metabolite and NMR spectroscopy experiments.

4.3.2. Purification of Metabolite

4.3.2.1. Small-scale purification trials

Before attempting large-scale fermentations, I carried out some smaller-scale trials in order to identify optimal conditions for the best production titre of the target metabolite. First, I grew cultures of *S. coelicolor* M1146-pCAPSalbC and *S. albus* J1074 in 10 mL Falcon™ tubes and took samples each day for a period of 7 days. I then carried out LC-MS analysis on the samples to determine how metabolite production changes over time (**Figure 47**). From this, it appeared that the ion intensity of m/z 324.16 ($[M+2H]^{2+}$) in *S. coelicolor* M1146-pCAPSalbC was the highest at day 4, after which the ion intensity of other pathway metabolites (m/z 510.27 and m/z 409.22) increase before diminishing after day 6. In *S. albus* J1074, overall production of m/z 324.16 ($[M+2H]^{2+}$) appears to be lower and peaks at a later time compared to *S. coelicolor* M1146-pCAPSalbC. I chose to focus on isolating the metabolite from *S. coelicolor* M1146-pCAPSalbC as production of the target metabolite appeared to be higher, and the lower metabolic background of the superhost should allow for easier separation of the metabolite from the crude microbial extract.

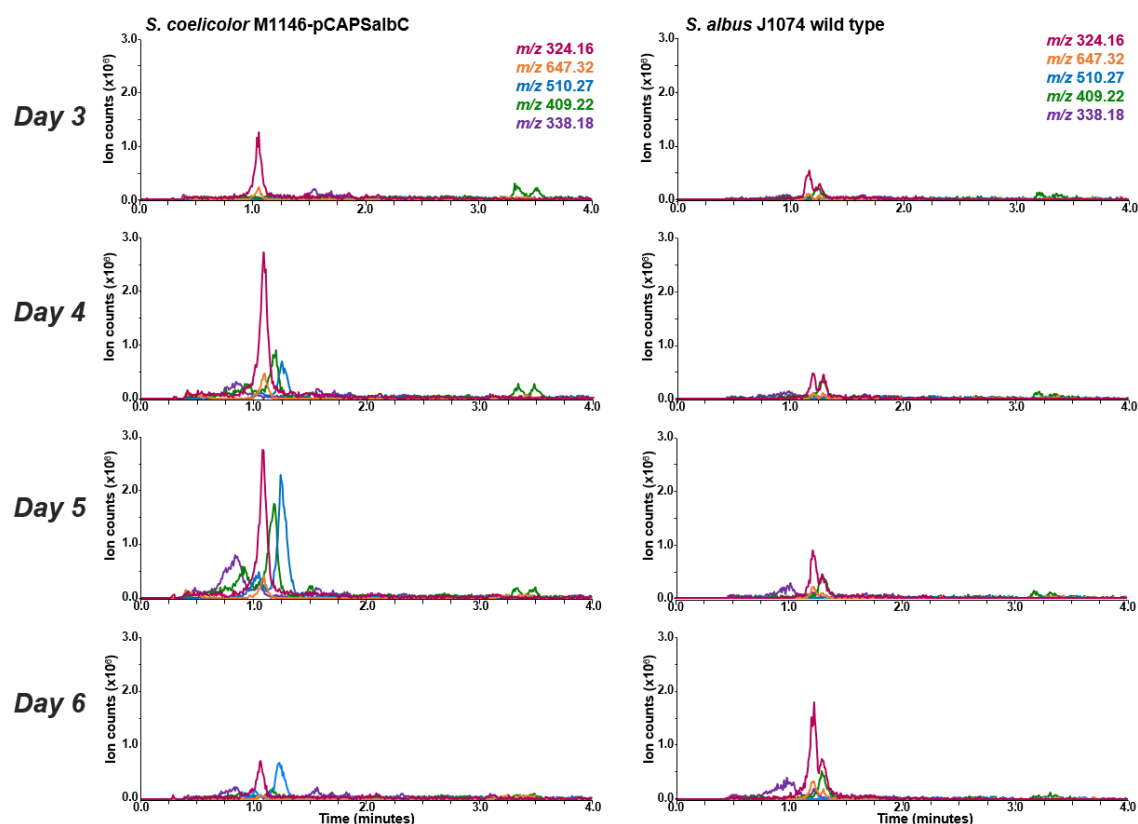


Figure 47: Time course of pathway-associated metabolite production. Production of metabolites over a 4-day period in *S. coelicolor* M1146-pCAPSalbC and *S. albus* J1074 is shown. Data obtained on a Shimadzu IT-TOF.

After this experiment, I trialled fermentations of *S. coelicolor* M1146-pCAPSalbC in a range of flask sizes (10 mL, 250 mL, 500 mL and 2 L) with different volumes of production media and analysed metabolite production at different time points. From this experiment, I found that metabolite production was only detected from 500 mL cultures grown in 2 L conical Duran flasks.

4.3.2.2. Large-scale purification

4.3.2.2.i. Liquid-liquid extraction

Following small-scale trials, I grew four batches of 500 mL SM12 cultures for a large-scale purification (methods described in section 6.5.1.). After four days of growth, the cultures were centrifuged and filtered, and the supernatant retained for purification. The fast elution of m/z 324.16 ($[M+2H]^{2+}$) in previous LC-MS analyses indicated that the metabolite is highly polar, therefore I washed the culture extract with ethyl acetate followed by butanol to remove hydrophobic material from the extract, retaining the target metabolite in the aqueous fraction. This aqueous fraction was dried down and retained for further purification steps.

4.3.2.2.ii. Solid phase extraction

The dried aqueous fraction was injected onto a SNAP HP20 Silica normal phase cartridge on a Biotage for separation using a gradient of increasing methanol. Fractions containing the target metabolite (**Appendix Figure 90**) were then combined and dried for further purification steps. The LC-MS chromatograms of extracts taken at each stage of purification is shown in **Figure 48**. This indicates that there were still background metabolites present in the sample following Biotage separation, therefore the extract was subjected to high-performance liquid chromatography (HPLC) purification.

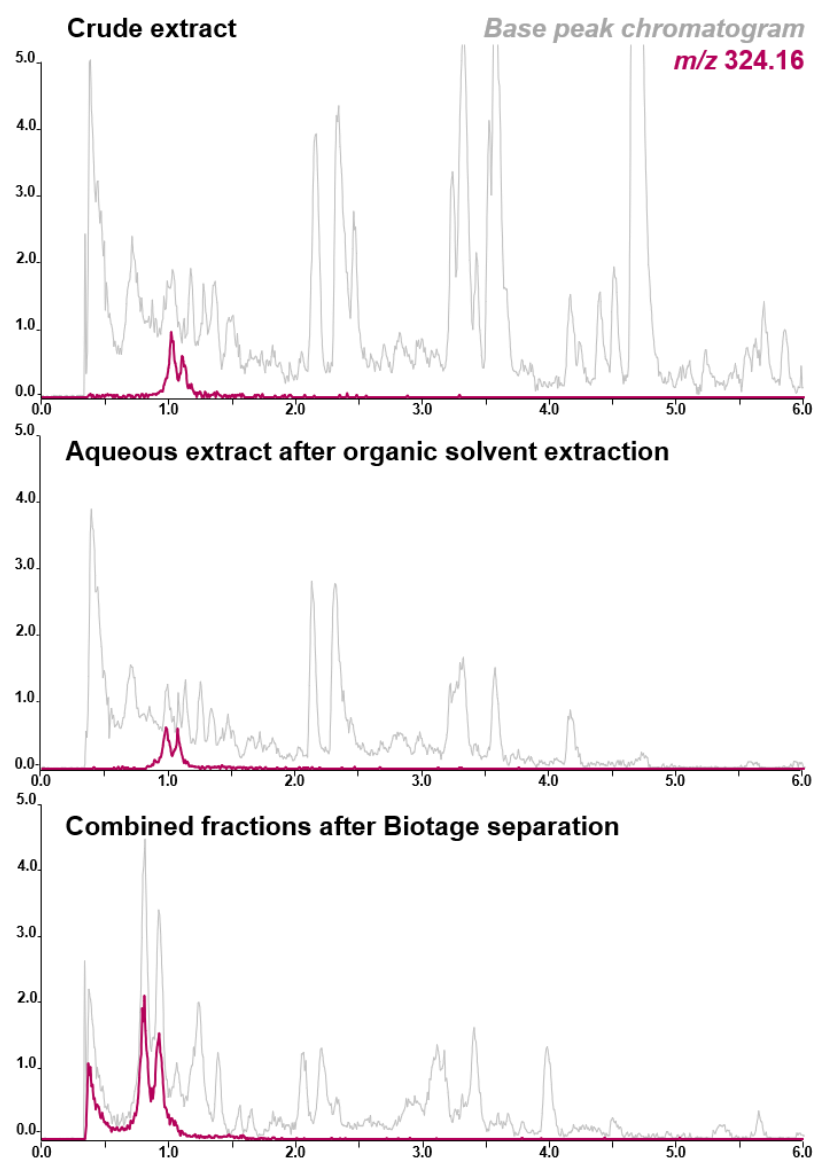


Figure 48: LC-MS analysis of fractions obtained during purification of the target pathway metabolite (m/z 324.16 $[M+2H]^{2+}$). LC-MS data obtained on a Shimadzu IT-TOF.

4.3.2.2.iii. High-performance liquid chromatography (HPLC)

The extract obtained following solid phase separation was first injected onto a PFP2 HPLC column, and fractions were manually collected when peaks appeared in the UV chromatogram (**Appendix Figure 91**). After analysing each fraction by LC-MS, I found that the peaks that appear between 12 and 15 minutes corresponded to elution of the target metabolite (**Figure 49**). Although the metabolite was almost pure at this stage, there was a small amount of background contamination as observed in the base peak chromatogram of the sample. Furthermore, the target metabolite appeared as two separate LC-MS peaks, and it was not clear whether these represented two different metabolites or structural isomers that might complicate the NMR analysis. It is also possible that the two LC-MS peaks represent different protonation states of the

metabolite. In order to investigate this, and to clean the sample as much as possible, I injected it onto a polar C18 HPLC column to see if the metabolites displayed better separation on a different column. I trialled a range of different organic solvent gradients with this column to see which one resulted in the best separation. From these trials, I found that applying an isocratic gradient of 10% methanol appeared to best separate the metabolites, based on the peak separation visible in the UV chromatogram at 210 nm (**Appendix Figure 92**). I manually collected fractions corresponding to each peak in the chromatogram, and LC-MS analysis indicated that I had purified the target metabolite that eluted as a single LC-MS peak (**Figure 50**). The total amount of pure compound obtained following these purification steps was 1.4 mg.

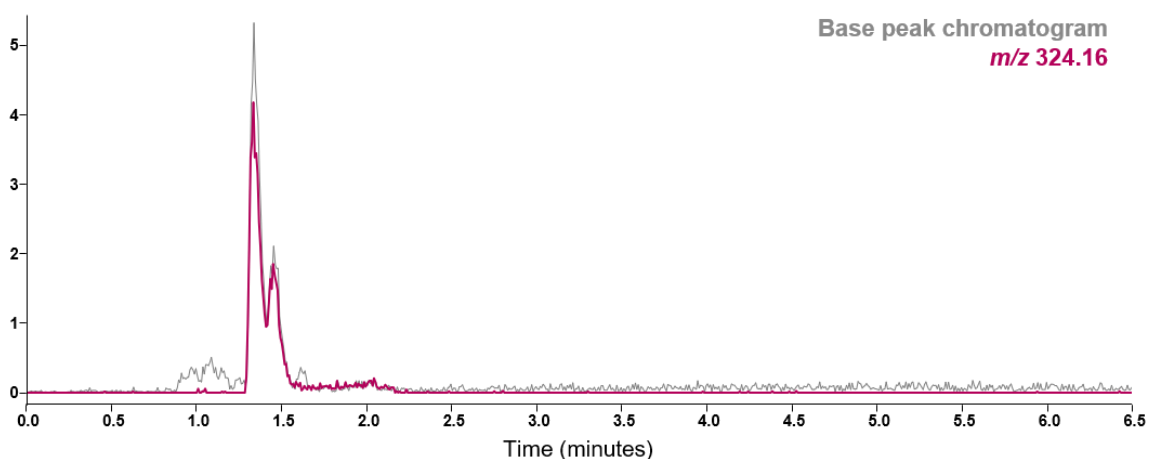


Figure 49: LC-MS analysis of combined HPLC fractions containing the target pathway metabolite (m/z 324.16 $[M+2H]^{2+}$). Obtained following separation of extract on a semi-preparative Phenomenex Luna PFP(2) column, with a 2-10% gradient of MeOH. Data obtained on a Shimadzu IT-TOF.

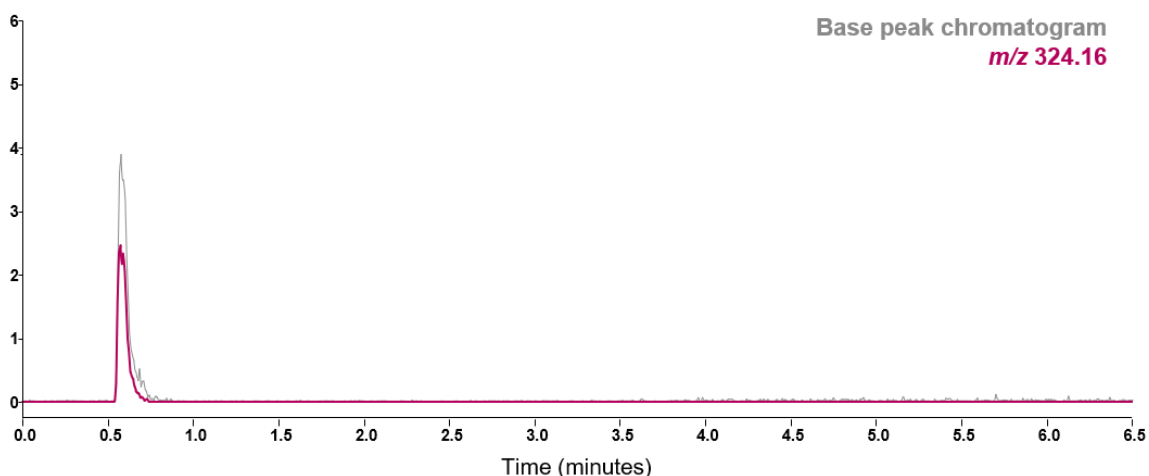


Figure 50: LC-MS analysis of pure metabolite obtained following a final HPLC purification step. Obtained following separation on a Luna Omega Polar C18 column with an isocratic gradient of 10% MeOH. Data obtained on a Shimadzu IT-TOF.

4.3.3. Structural Elucidation

4.3.3.1. Nuclear magnetic resonance (NMR) analysis

In order to elucidate the structure of the newly isolated RiPP, I carried out a series of NMR experiments (methods described in section 6.5.2.1.). Purified metabolite was dissolved in ~600 μL of DMSO-d_6 and subjected to a range of 1D and 2D NMR experiments including ^1H , ^{13}C , correlation spectroscopy (COSY), heteronuclear single quantum correlation edited (HSQCed), heteronuclear multiple bond correlation (HMBC), total correlation spectroscopy (TOCSY), nuclear overhauser effect spectroscopy (NOESY) and HSQC-TOCSY. Detailed correlation data are shown in **Figure 52**, 1D NMR spectra are shown in **Figures 54** and **55** and 2D NMR spectra are shown in **Appendix Figures 99-102**. A list of chemical shifts is shown in **Table 2**. The data obtained from NMR allowed me to piece together the structure of this metabolite, revealing that it does derive from the previously predicted HLSATH core peptide.

4.2.2.1.i. Elucidating the peptide backbone

The presence of the histidine residues was indicated through carbons with characteristic ^{13}C shifts of δ_{C} 118 ppm, δ_{C} 122 ppm and δ_{C} 134 ppm, which are similar to those in the imidazole rings of characterised histidine residues (356). The quaternary carbon of each ring was difficult to see via 1D ^{13}C NMR, but 2D data revealed the presence of these carbons with characteristic chemical shifts (δ_{C} 132 ppm and δ_{C} 135 ppm). Adjacent residues could then be pieced together based on 2D correlations. Long-range HMBC correlations from each carbonyl connected atoms within each residue as well as with neighbouring residues. NOESY correlations, which connect atoms that are close together spatially, helped to connect amide protons with the CH of neighbouring residues. TOCSY correlations indicate which protons belong to a single spin system, which could therefore connect together all protons within each amino acid residue. These connections could be further confirmed through COSY correlations, which connect protons attached to neighbouring carbon atoms. HSQC-TOCSY correlations also confirmed which carbons and protons were present within a single residue. HSQC correlations confirmed which protons were attached to each carbon within the metabolite.

4.2.2.1.ii. Elucidating the post-translational structural modification

Although there were characteristic ^{13}C and ^1H shifts and correlations indicating the presence of the HLSATH peptide backbone, the exact identity of the post-translational modification was difficult to confirm. Although I initially predicted that there could be an oxazoline or oxazole ring in the metabolite, the chemical shifts of the serine side chain indicated that this residue was unmodified. Interestingly, the ^{13}C shift for the sp^2 carbon between Leu2 and Ser3 was δ_{C} 157.1 ppm, which differed from either an unmodified

amide carbonyl (expected $\delta_C \sim 170$ ppm) or an oxazoline ring (expected $\delta_C \sim 140$ ppm). Instead, this carbon chemical shift was similar to that of the corresponding carbons in the amidine rings of bottromycin (δ_C 157.9 ppm in $CDCl_3$) (357) and klebsazolicin (δ_C 156.8 ppm in $DMSO-d_6$) (129), suggesting that an amidine ring might be present in the metabolite. Furthermore, the terminal NH_2 group that would be present on an unmodified N-terminal histidine could not be seen by NMR. Overall, this suggested that an amidine ring was present between the N-terminal amine of His1 and the carbonyl of Leu2 (core peptide numbering). The NMR correlation data supported the presence of a six-membered amidine ring, including HMBC correlations between C15-H9 (δ_H 8.05 ppm) and C15-H18 (δ_H 4.30-4.27 ppm) (**Figure 53**). A particularly diagnostic HMBC correlation could be seen between C15-H7 (δ_H 3.95-3.91 ppm), which would not fit an oxazoline-containing structure due to the distance between these two atoms (**Figure 51**). An exocyclic double bond between C15 and N17 was inferred as NMR data did not reveal an amide proton attached to N17.

Based on the widespread occurrence of homologous biosynthetic gene clusters in streptomycetes, and the presence of an amidine ring in the structure, I named this metabolite streptamidine. To reflect this, the genes within the biosynthetic pathway were also named according to the acronym *ami*. The discovery of a novel amidine-containing metabolite is particularly exciting as it is a rare structural feature to be found in natural products. To date, only two examples of amidine-containing RiPPs are known: klebsazolicin, which contains a six-membered amidine ring (129) and bottromycin, which contains a 12-membered macroamidine ring (146) (**Figure 64**).

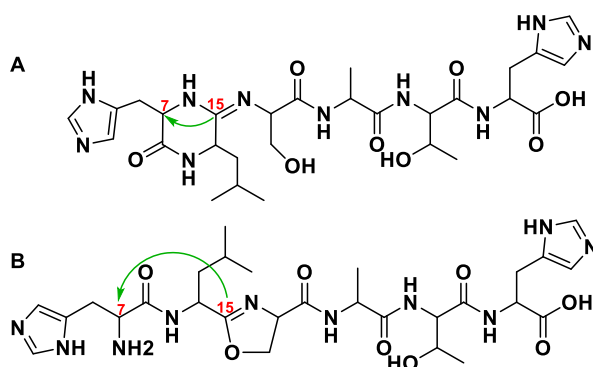


Figure 51: A: Diagnostic HMBC correlation for the amidine ring in streptamidine. The correlation between C15 (δ_C 157.9 ppm) and H7 (δ_H 3.95-3.91 ppm) is too long a distance to support the presence of an oxazoline-containing structure (B).

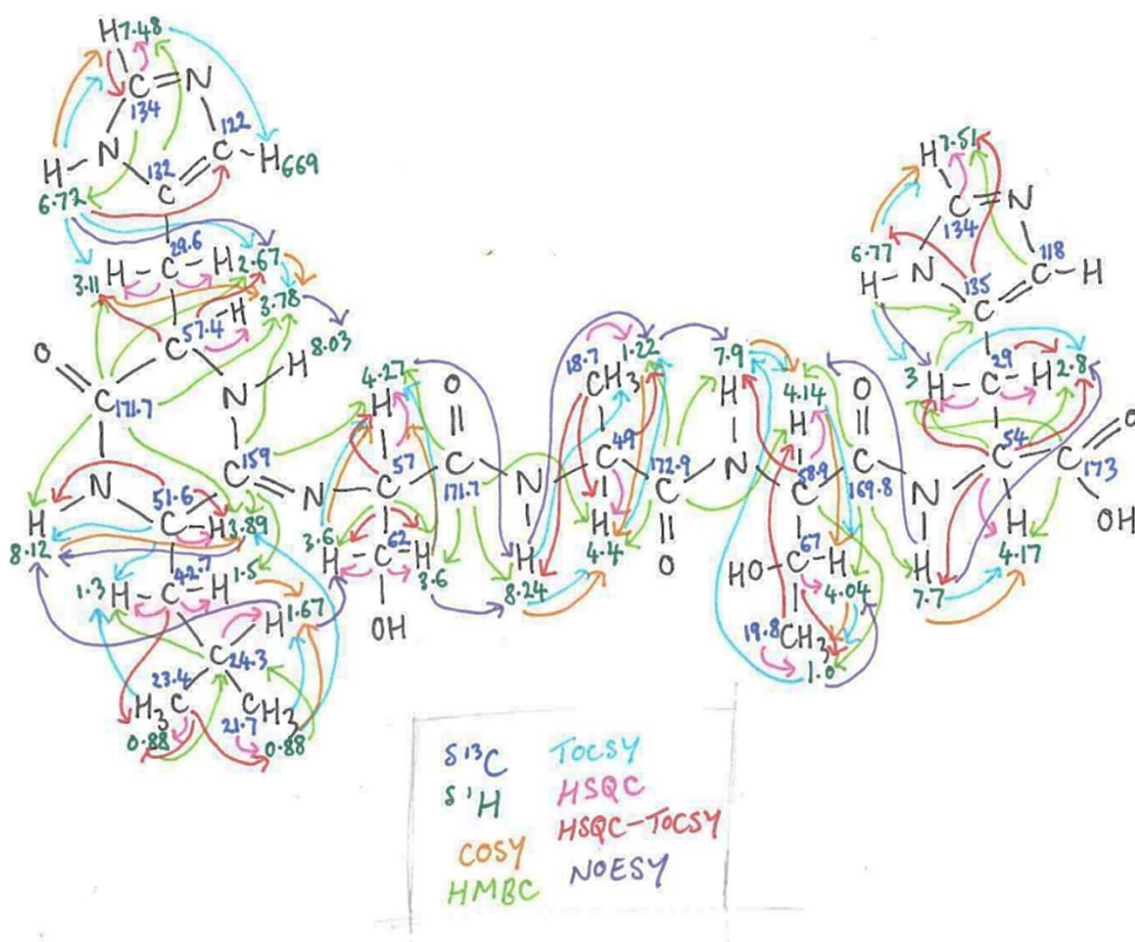


Figure 52: Detailed correlation data for the structural elucidation of streptamidine.

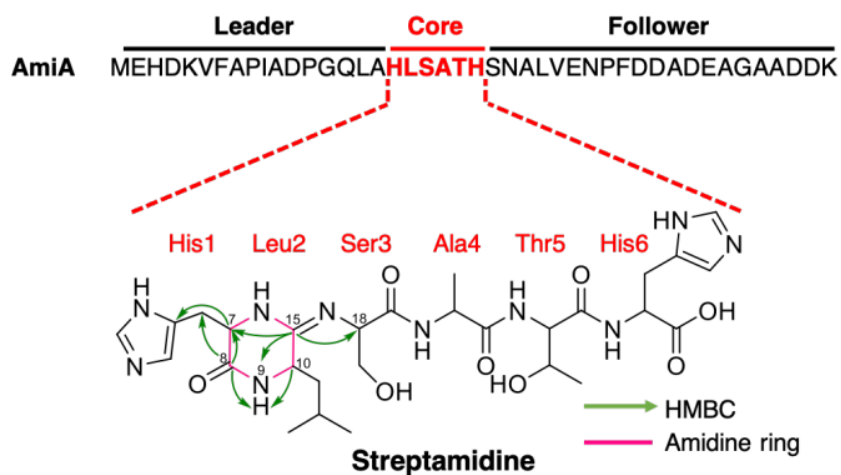


Figure 53: Structure of streptamidine in relation to the precursor peptide AmiA. Leader, core and follower regions of the precursor peptide are highlighted along with key HMBC correlations that support the presence of the amidine ring in the metabolite.

Table 2: NMR chemical shift assignments for streptamidine in DMSO-d₆

C/N number	Amino acid	C chemical shift δ	H chemical shift δ	Multiplicity	Coupling Constant in Hz
1	His1	--	--	--	
2	His1	134.7	7.48	s	
3	His1	--			
4	His1	131.6 ^a	--	--	
5	His1	121.0 ^b	6.67	s	
6	His1	32.6	2.84, 2.75	dd, dd	$J_{6a,6b} = 14.1$, $J_{6a,7} = 5.2$, $J_{6b,7} = 7.4$
7	His1	58.8	3.95-3.91	m	
8	His1	170.7	--	--	
9	Leu2	--	8.05	s	
10	Leu2	51.0	3.82	d	$J_{10,11a} = 10.8$
11	Leu2	45.9	1.41-1.35, 0.77-0.71	m, m	
12	Leu2	23.8	1.64-1.57	m	
13	Leu2	21.4	0.82	d	$J_{12,13} = 6.4$
14	Leu2	24.1	0.78	d	$J_{12,14} = 6.6$
15	Leu2	157.1	--	--	
16	His1	--			
17	Ser3	--	--	--	
18	Ser3	57.1	4.30-4.27	m	
19	Ser3	62.2	3.62	d	$J_{18,19} = 5.9$
20	Ser3	171.4	--	--	
21	Ala4	--	8.09	d	$J_{21,22} = 7.0$
22	Ala4	48.9	4.41-4.35	m	
23	Ala4	18.7	1.25	d	$J_{22,23} = 7.1$
24	Ala4	172.9	--	--	
25	Thr5	--	7.88	d	$J_{25,26} = 8.6$
26	Thr5	59.3	4.14	dd	$J_{25,26} = 8.6$, $J_{26,27} = 3.6$
27	Thr5	66.9	4.02-3.98	m	
28	Thr5	19.9	1.00	d	$J_{27,28} = 6.3$
29	Thr5	170.0	--	--	
30	His6	--	7.78	d	$J_{30,31} = 7.5$
31	His6	53.5	4.27-4.24	m	
32	His6	29.2	2.97, 2.82	dd, dd	$J_{32a,32b} = 14.8$, $J_{31,32a} = 5.1$, $J_{31,32b} = 7.4$
33	His6	134.1 ^a	--	--	
34	His6	117.1 ^b	6.81	s	
35	His6	--	--	--	
36	His6	135.1	7.54	s	
37	His6	--		--	
38	His6	173.1	--	--	

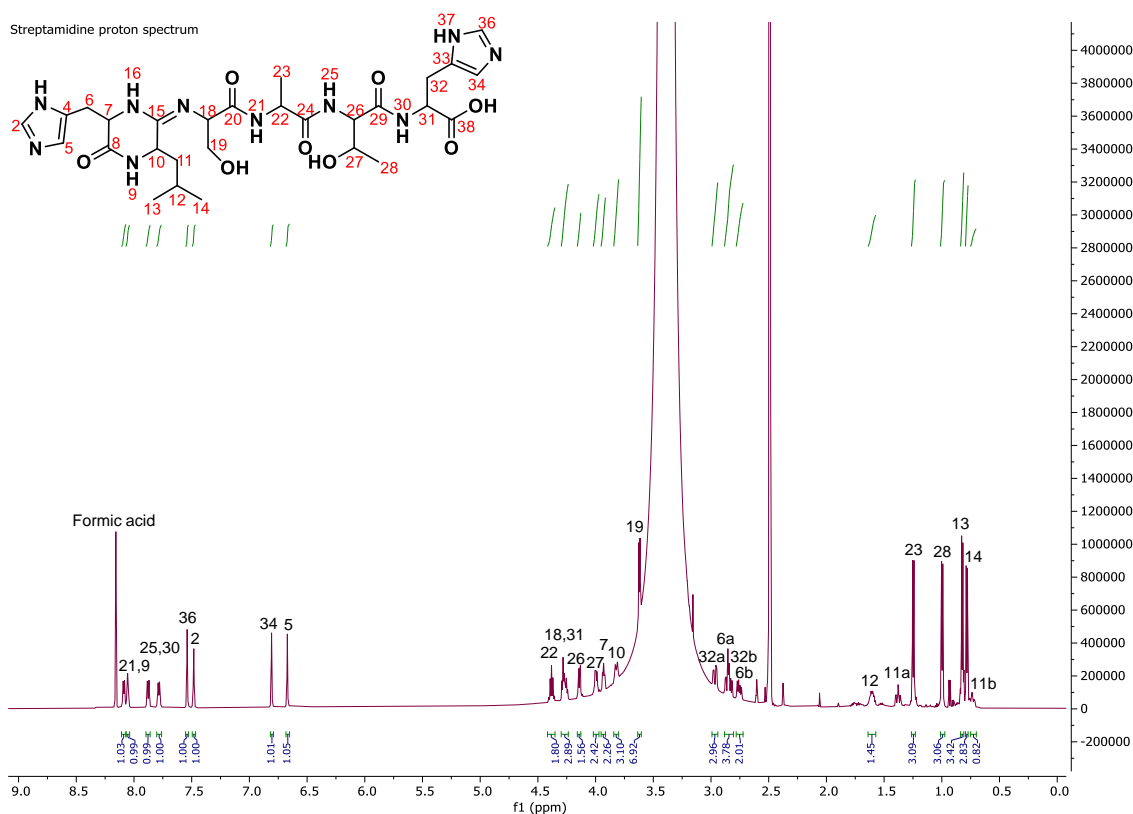


Figure 54: Full proton spectrum for streptamidine recorded in DMSO- d_6 . Chemical shifts for each proton in streptamidine are indicated, along with the integration of each peak. Data obtained on a Bruker Ascend 600 MHz instrument.

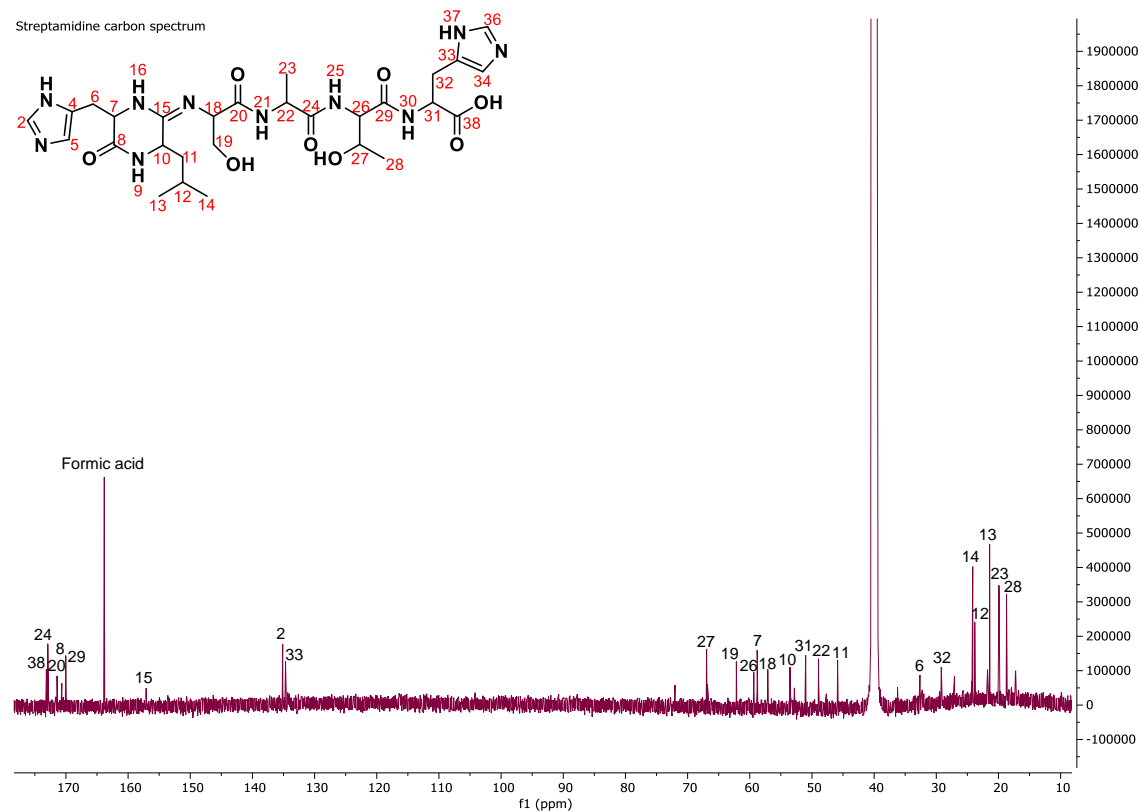


Figure 55: Full carbon spectrum for streptamidine in DMSO- d_6 with numbered atoms. Data obtained on a Bruker Ascend 600 MHz instrument.

4.3.3.2. Structural confirmation using a fluorescamine-binding assay

Although the NMR data clearly supported the presence of an amidine ring in streptamidine, I was interested to prove this further by testing for the presence or absence of the terminal amine group. To achieve this, I carried out a binding assay of fluorescamine (4-phenylspiro-[furan-2(3H),1-phthalan]-3,3'-dione) with purified streptamidine (method described in section 6.5.2.2.). Fluorescamine is a reagent that reacts with free primary amine groups to produce a fluorescent complex. If the proposed structure of streptamidine is correct, the metabolite should not react with fluorescamine as it does not contain a free primary amine group. If the metabolite contained an oxazole or oxazoline however, there would be a primary amine group free to form a fluorescamine-bound complex. As a control, I mixed unmodified L-histidine with fluorescamine, which formed the expected complex as indicated by LC-MS analysis (m/z 416.12). The multiple peaks observed could be a consequence of multiple protonation states or stereoisomers. However, streptamidine did not react with fluorescamine to form any observable complex by LC-MS, or any fluorescence detected by UV (**Figure 56**). This experiment therefore supports the amidine-containing structure elucidated from NMR experiments.

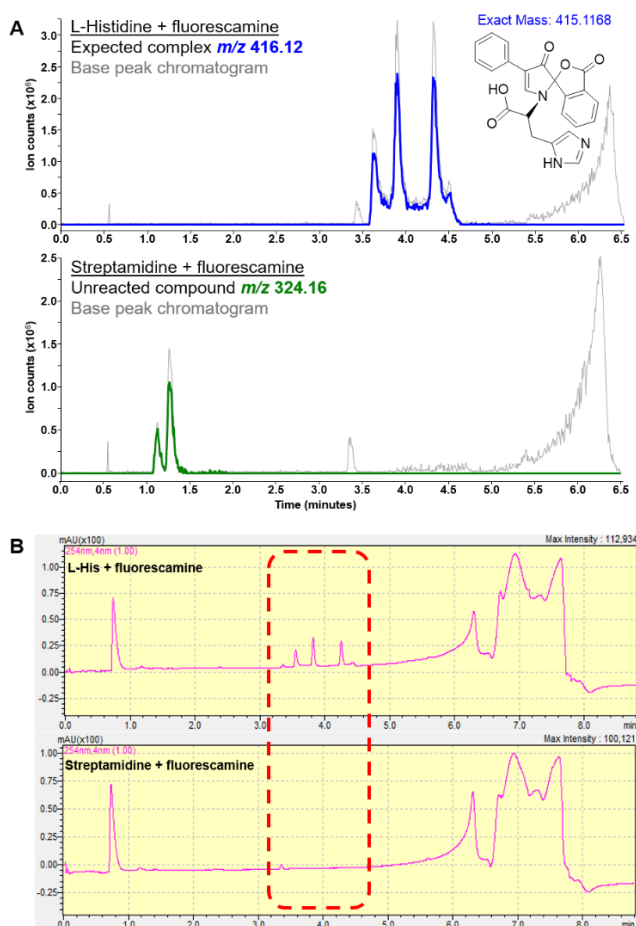


Figure 56: A: LC-MS analysis of fluorescamine reaction with L-histidine and streptamidine. B: Corresponding UV trace of samples obtained with the PDA detector. Data obtained on a Shimadzu IT-TOF.

4.3.3.3. Determination of stereochemistry of streptamidine

Once the structure of streptamidine was confirmed, I was interested to determine the exact stereochemical configuration of the metabolite. To achieve this, I carried out a Marfey's analysis on a sample of the pure metabolite (methods described in section 6.5.2.3.). This involved hydrolysis of streptamidine into its individual amino acid components, followed by derivatisation with Marfey's reagent L-FDAA (N_{α} -(2,4-dinitro-5-fluorophenyl)-L-alaninamide). The derivatised components could then be compared with L- and D- amino acid standards derivatised with the same reagent. As streptamidine derives from a ribosomal pathway, and there are no epimerases in the biosynthetic gene cluster, it would be expected that the component amino acids are present in L- form. The results of the Marfey's analysis revealed that histidine, serine and threonine were indeed present in L- form. Alanine was also present in L- form, although there were traces of D- alanine in the hydrolysed streptamidine sample. Interestingly however, leucine was detected in equal amounts of both L- and D- configuration (**Figure 57**), suggesting that streptamidine might exist in two isoforms.

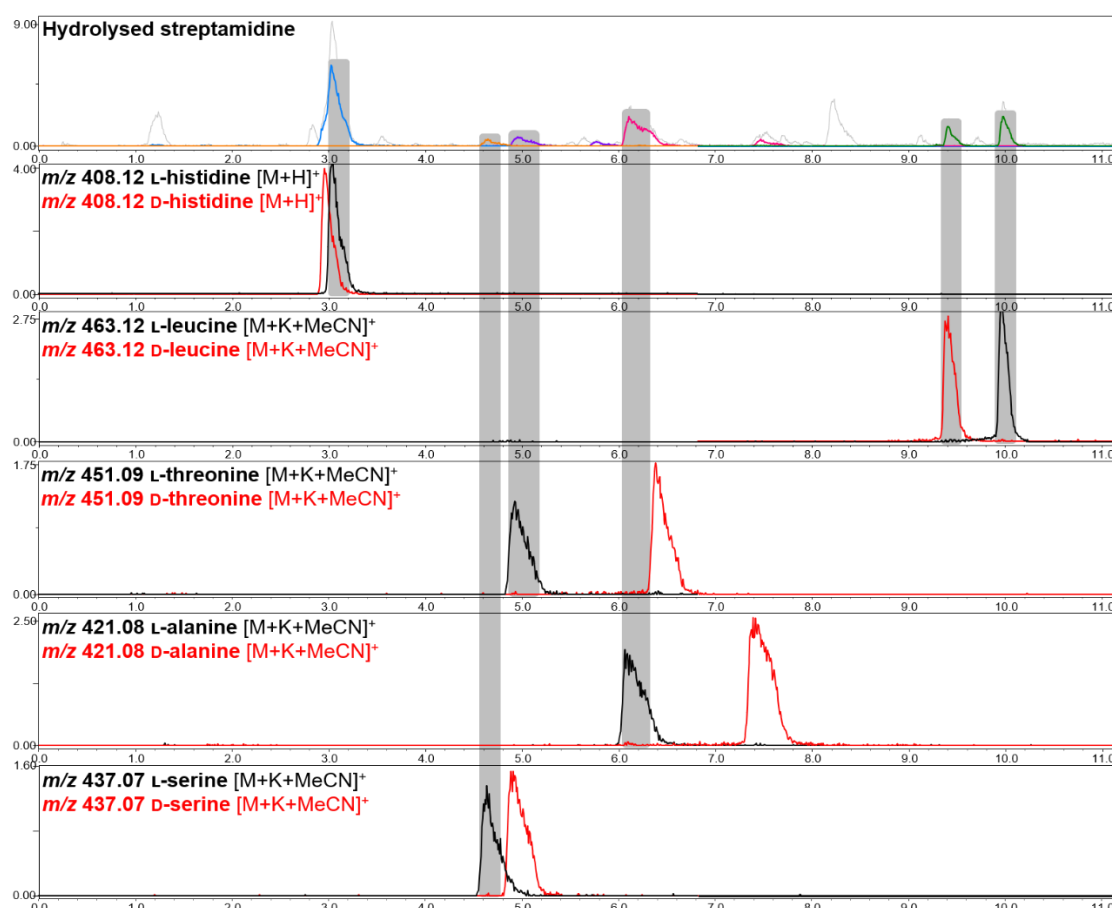


Figure 57: LC-MS analysis of Marfey's analysis of streptamidine. The top panel shows the hydrolysed streptamidine sample where constituent amino acids have been derivatised with Marfey's reagent. Below this are the LC-MS chromatograms obtained from derivatisation of individual amino acid standards of L- and D- amino acids.

The Marfey's analysis was complicated by the observation of unusual adducts via LC-MS: although histidine was detected as a proton adduct ($[M+H]^+$, m/z 408.12), the other amino acids were detected as potassium and acetonitrile adducts ($[M+K+MeCN]^+$) in both the amino acid standards and the hydrolysed sample. The structure of streptamidine with all amino acids in L- configuration is shown in **Figure 58**.

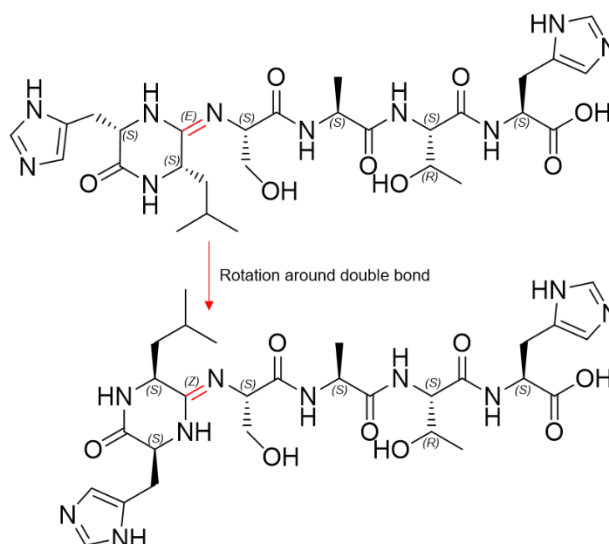


Figure 58: Stereochemistry of streptamidine. Structures show configuration of streptamidine with L-amino acids.

4.3.3.4. Spontaneous isomerisation of streptamidine

The result of the Marfey's analysis suggests that streptamidine exists as two isoforms, with epimerisation occurring on the leucine residue. This could explain why streptamidine was observed as two separate peaks by LC-MS analysis during previous metabolomic analysis (e.g. **Figure 49**). Interestingly, when streptamidine was analysed by NMR, additional signals could be observed in the spectra during the time period that NMR experiments were being carried out. A comparison of the proton spectra obtained before and after the long 2D experiments were carried out shows that additional proton signals are present in the second spectrum (**Figure 59**). This indicates that streptamidine might be spontaneously isomerising over time. Furthermore, the additional NMR signals are observed within the region of the leucine residue and amidine ring, supporting the idea that leucine might be epimerising.

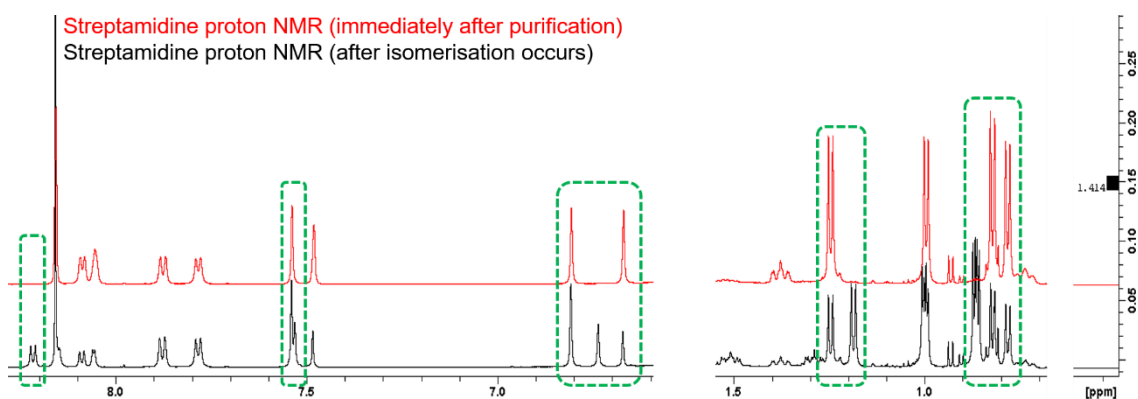


Figure 59: Comparison of ^1H NMR spectra before and after isomerisation of streptamidine. The red spectrum shows proton signals of the freshly purified metabolite, whereas the black spectrum shows proton signals after the metabolite has isomerised over time. Areas of the spectrum where additional proton signals are present are highlighted in green boxes.

4.3.3.5. MicroED analysis of streptamidine

In order to finally confirm the structural configuration of streptamidine, a “mixed isomer” sample of streptamidine has been submitted for a preliminary MicroED (microcrystal electron diffraction) (348,349) analysis to be completed in November 2020 at the Diamond Light Source. Due to this approach of capturing single crystals of the metabolite on a grid, it is possible that this analysis may allow visualisation of one or both of the proposed streptamidine isomer structures, allowing final confirmation of the structure of streptamidine.

4.3.3.6. Structures of streptamidine-related pathway metabolites

As well as streptamidine, several other pathway-related metabolites were identified during the metabolomic screening described in Chapter 3. Once the structure of streptamidine had been elucidated, I could then look back at the other metabolites identified to see how they might be structurally related. In Chapter 3 (**Figure 36**), I showed that the four pathway-related metabolites (m/z 647.32, m/z 510.27, m/z 409.22 and 338.18) all appear to be related based on similar fragmentation patterns and core peptide predictions. As such, the identified masses are likely to represent amidine-containing peptides derived from HLSAT (m/z 510.27), HLSA (m/z 409.22) and HLS (m/z 338.18) (**Figure 60**). The predicted structures of the HLSAT- and HLSA-derived metabolites are also supported by accurate mass data for the metabolites obtained on a Synapt G2Si (calculated m/z 510.2671, observed m/z 510.2668; calculated m/z 409.2194, observed m/z 409.2195; **Appendix Table 26**). The common MS^2 fragment ions observed from all four metabolites are annotated in **Figure 61**, which also support the proposed structures of these metabolites. The HLS-derived metabolite (m/z 338.18)

was not detected in sufficient intensity by high resolution LC-MS to obtain accurate mass and fragmentation data.

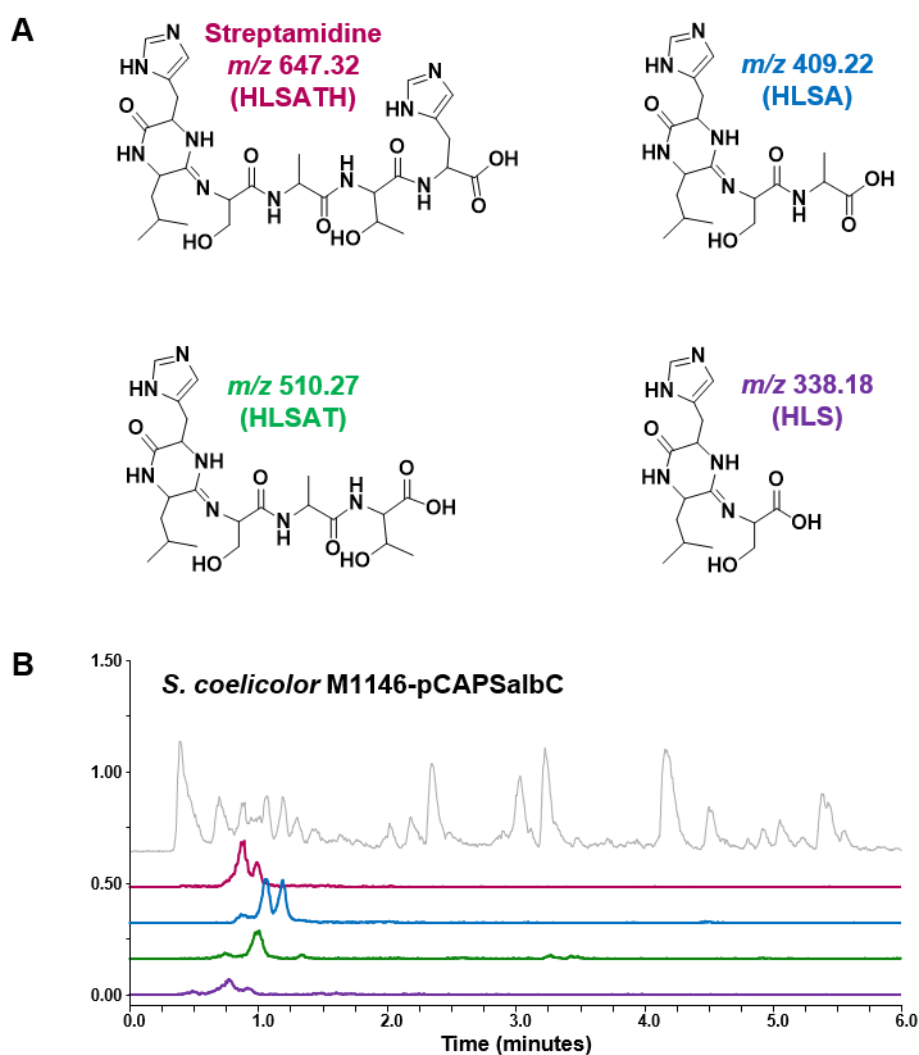


Figure 60: A; Predicted structures of streptamidine-related metabolites. Structures were inferred from the elucidated structure of streptamidine as well as high resolution LC-MS² data for the related metabolites. B: LC-MS data showing extracted ion chromatograms of the corresponding metabolites, obtained on a Shimadzu IT-TOF.

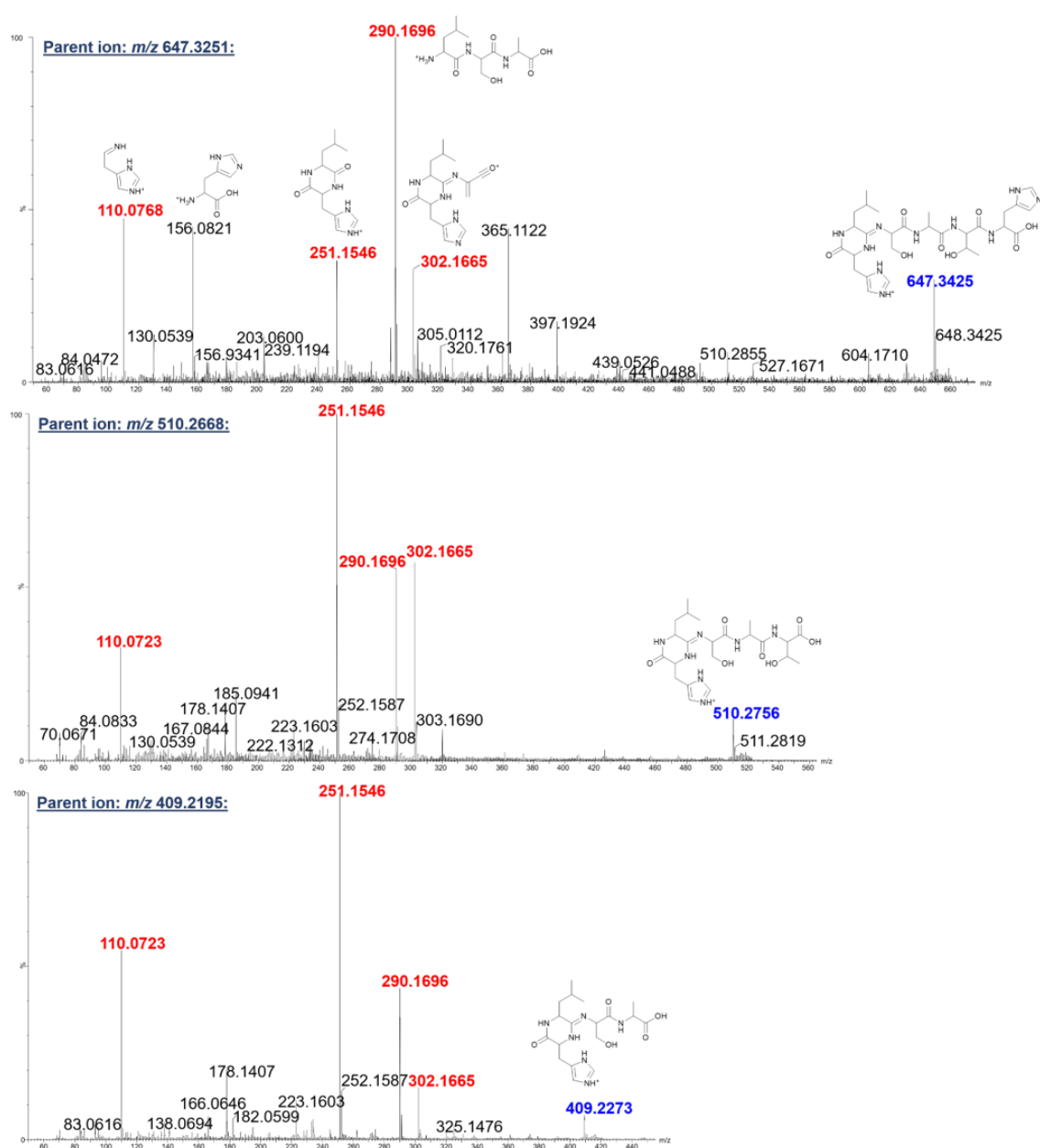


Figure 61: High-resolution LC-MS² data for putative pathway-related metabolites. Data for m/z 647.32, m/z 510.27 and m/z 409.22, obtained using a Waters Synapt G2Si. Common fragment ions are highlighted in red and annotated with putative fragment structures.

4.3.3.7. Is streptamidine the final pathway product?

The pathway-related metabolites identified and described above were repeatedly detected by LC-MS following several independent fermentations of *S. coelicolor* M1146-pCAPS**albC** and *S. albus* J1074. Although the largest metabolite identified was that of streptamidine (singly charged m/z 647.32), this hexapeptide represents a small core region in between both a follower and leader region of the precursor peptide. Although there are examples of RiPPs containing both follower and leader regions, we wondered whether there could be a larger pathway-related metabolite being produced that was not being detected during my LC-MS experiments. With this in mind, LC-MS samples of *S. coelicolor* M1146-pCAPS**albC** and *S. coelicolor* M1146-pCAPS**albC** Δ *amiA* extracts were analysed by other mass spectrometers optimised for detection of larger peptides and proteins. This included a MALDI-TOF analysis (carried out by Carlo de Oliveira Martins) and a Synapt G2Si analysis followed by a comparative metabolomics analysis using Progenesis (carried out by Gerhard Saalbach). The results of these additional mass spectrometry analyses did not reveal any additional larger masses that might be indicative of a larger final product of the biosynthetic gene cluster.

4.3.4. Investigating the Biosynthesis of Streptamidine

Following structural confirmation of streptamidine, I was interested to investigate the biosynthesis of the metabolite, and how this relates to the RiPP tailoring enzymes present in the biosynthetic pathway. As amidine rings are rare structural features, there is limited information about how these are biosynthesised in RiPPs. Furthermore, with the elucidation of a single post-translational modification, the roles of individual RiPP tailoring enzymes within the streptamidine pathway would be interesting to investigate.

4.3.4.1. Biosynthetic shunt metabolites

As discussed in Chapter 3, deletion of the dehydrogenase gene (*amiE*) led to increased production of three particular metabolites: m/z 272.16 (peptide A), m/z 314.17 (peptide B) and m/z 332.18 (peptide C) (**Figure 62**). As this gene deletion could be complemented, these metabolites were not produced as a result of polar effects of the deletion. I therefore hypothesised that these peptides could provide clues about the biosynthetic mechanism of amidine ring formation. Peptides A and B share a common MS² fragment (141.01) observed by high-resolution LC-MS analysis (**Appendix Figure 103**), suggesting that these metabolites are structurally related. The exact mass of peptide C (calculated m/z 332.1816, observed m/z 332.1818) indicated that it could correspond to an acetylated Leu-Ser-Ala tripeptide, which would be consistent with a region within the streptamidine core peptide. To confirm this, I obtained a synthetic standard of an N-acetylated LSA tripeptide, and compared its LC-MS trace to that of an extract of *S. coelicolor*-pCAPS Δ *amiE*. I also spiked the Δ *amiE* extract with the synthetic standard. The co-elution of peptide C at the same retention time in all three samples (observed as a sodium adduct m/z 354.16), indicated that this metabolite does indeed correspond to the LSA tripeptide (**Figure 63**). Based on this, I hypothesised putative structures for peptides A and B, whose masses could theoretically correspond to a dehydrated LSA peptide (peptide A, m/z 272.16) and an acetylated dehydrated LSA tripeptide (peptide B, m/z 314.17). I therefore proposed that the dehydration of this peptide could correspond to oxazoline heterocycle formation on the central serine residue (**Figure 62**).

4.3.4.2. Biosynthetic mechanism of amidine formation

The production of putative oxazoline-containing intermediates could suggest that an intermediate ring structure is formed prior to ultimate formation of the amidine ring. Interestingly, this parallels one of the possible biosynthetic mechanisms proposed by Travin *et al* for amidine ring formation in klebsazolicin (**Figure 64**), which occurs between Ser1 and Gln2 of the KlpA precursor peptide (205). While attempting to dissect amidine biosynthesis by site-specific mutagenesis of the precursor, Travin *et al* noticed an important role of the hydroxyl group on the side chain of Ser3: if this residue was mutated to any amino acid other than threonine, the amidine cycle is not produced. A biosynthetic mechanism was therefore suggested involving formation of an intermediary structure following attack of the side chain of Ser3 on the Gln2 carbonyl, before immediate attack from the N-terminal amine group causing rearrangement and ultimate formation of the amidine ring (205). Based on this, I proposed a similar biosynthetic mechanism for streptamidine (**Figure 65**).

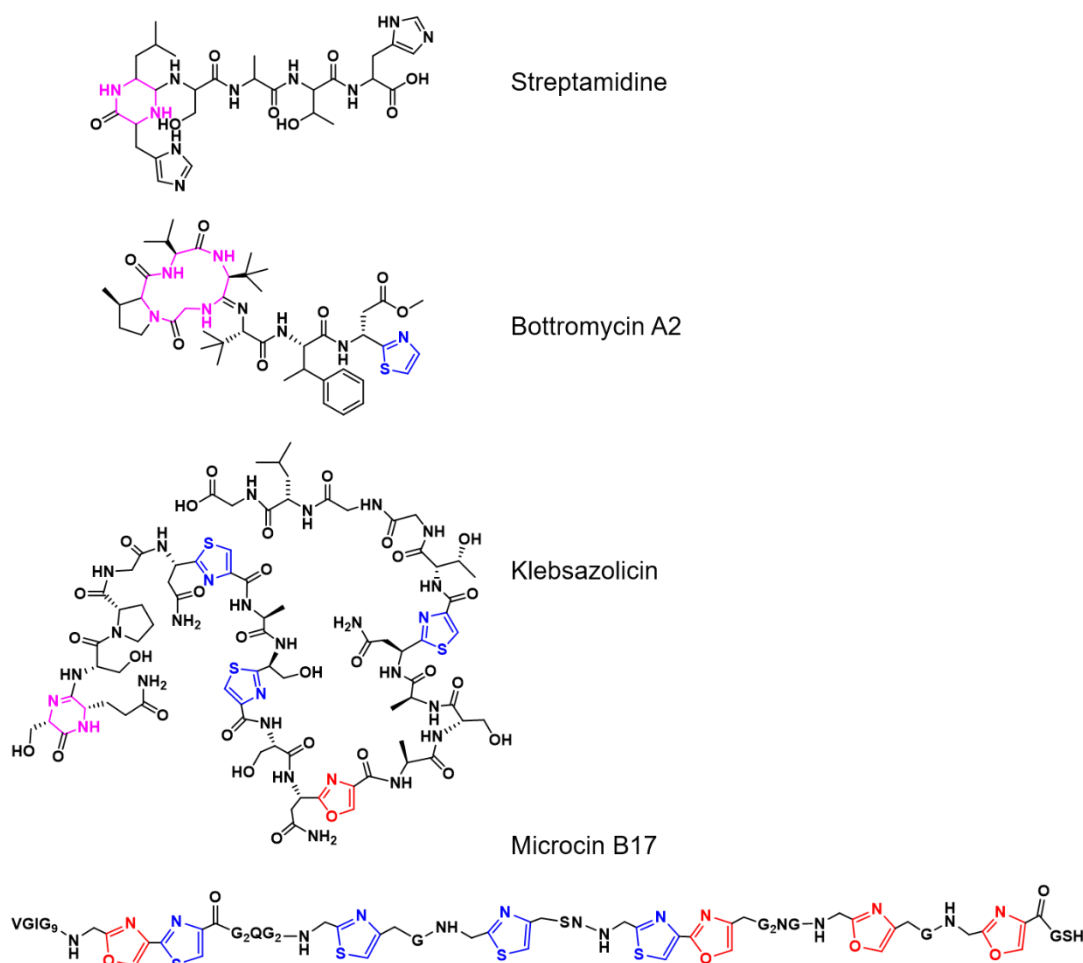


Figure 64: Structures of RiPPs containing amidine rings (pink), thiazole heterocycles (blue) and oxazole heterocycles (red).

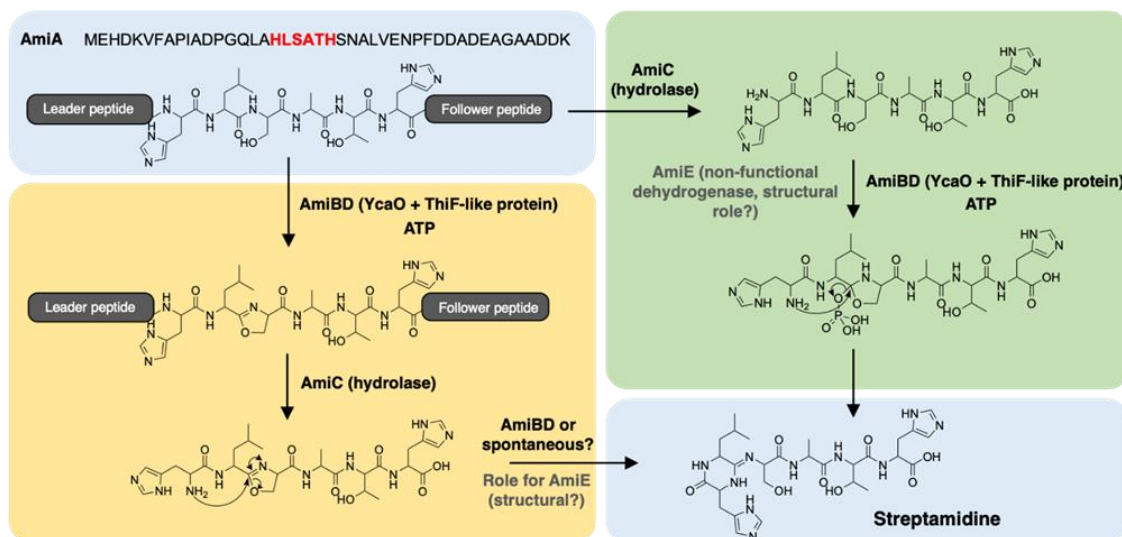


Figure 65: Proposed mechanism of amidine ring formation in streptamidine. Based on predicted intermediates produced by dehydrogenase mutant. The yellow box represents a route that goes via a stable oxazoline intermediate prior to leader/follower peptide removal, where AmiE potentially has a structural role within an AmiBDE complex for proper amidine formation but is not needed for initial ATP-dependent oxazoline formation. The green box represents a route where the leader/follower peptide is removed first, thereby providing a free N-terminal amine for cyclisation via an O-phosphorylated hemiorthoamide. In the absence of AmiE, the final amidine forming step could potentially be disrupted.

4.2.4.3. Dissecting the roles of RiPP tailoring enzymes in streptamidine biosynthesis

The gene deletion experiments described in Chapter 3 suggested that four RiPP tailoring enzymes in the streptamidine biosynthetic pathway were essential for biosynthesis of the metabolite, as deletion of their genes abolished production. These were the E1-like protein (AmiB), the hydrolase (AmiC), the YcaO-domain protein (AmiD) and the dehydrogenase (AmiE) (Chapter 3, **Figure 42** Figure 43).

4.2.3.3.i. Role of YcaO-domain protein AmiD and E1-like protein AmiB

In klebsazolicin biosynthesis, a YcaO-domain protein catalyses amidine formation with cooperation from the E1-like partner protein. I therefore propose that AmiB and AmiD are responsible for amidine formation in streptamidine (**Figure 65**). Interestingly, YcaO-domain proteins have been shown to be both mono- and bi-functional in their catalytic activity. In klebsazolicin biosynthesis, a single YcaO-domain protein (KlpD) catalyses both amidine ring and oxazoline heterocycle formation (205). In bottromycin biosynthesis, two distinct YcaO-domain proteins separately catalyse thiazoline formation (BtmE) and macroamidine formation (BtmF) (204,358). In streptamidine, an amidine ring is formed, but the presence of an unmodified threonine residue in the metabolite suggests that azoline heterocycle formation is not catalysed by AmiD. This protein might therefore be a mono-functional enzyme responsible only for amidine ring formation. The

amino acid sequence of YcaO-domain proteins can also give further clues about their catalytic activity. For example, Dunbar *et al* showed that azoline-forming YcaO-domain proteins contain a conserved PxPxP motif at the C-terminus. Truncation of these residues, or mutation of adjacent residues, led to abolished catalytic activity of BalhD, an azoline-forming YcaO from *Bacillus* sp. A1 (196). The C-terminus of the streptamidine YcaO protein AmiD has a sequence of PAPHM, which is similar but does not match the precise motif reported to be present in azoline-forming proteins. An alignment and MEME analysis of all the YcaO-domain proteins associated with the streptamidine-like gene clusters identified in Chapter 2 shows that the majority of these sequences have a conserved proline-rich LAPPPHM motif at the C-terminus (**Figure 66**). This suggests that a proline-rich region might also be important for amidine ring catalysis.

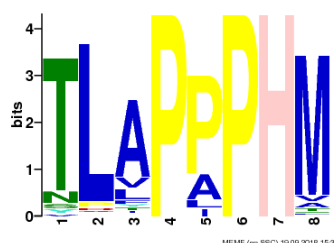
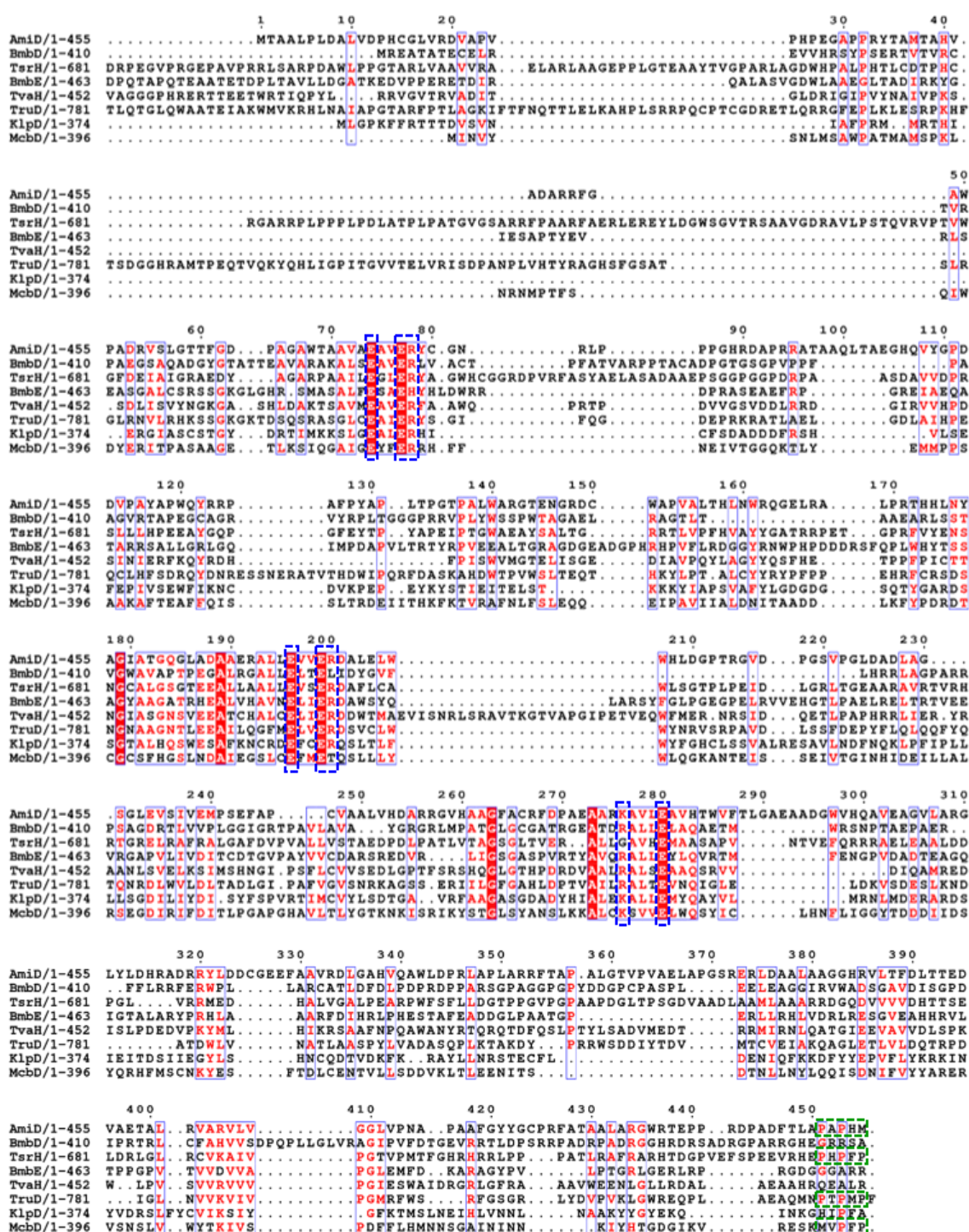


Figure 66: MEME analysis of all streptamidine-like YcaO-domain proteins. The eight final residues at the C-terminus of each YcaO-domain protein associated with streptamidine-like precursor peptides were submitted for analysis to determine whether a proline-rich sequence was abundant, which is characteristic of azole-forming YcaO proteins.

An alignment of YcaO protein sequences from several different RiPP pathways is shown in **Figure 67**, including examples of proteins involved in catalysis of azoline formation (KlpD, BmbD, TsrH, TruD and McbD), amidine formation (AmiD, KlpD and BmbE) and thioamide formation (TvaH). The key catalytic residues of YcaO-domain proteins are conserved across these sequences. As highlighted on the alignment figure, the C-terminal PxPxP motif identified by Dunbar *et al* is present in azoline-forming YcaO proteins such as TsrH and TruD, but absent from the bi-functional klebsazolicin YcaO (KlpD) and the azoline-forming YcaO from the *S. bottropensis* bottromycin pathway (BmbD). This suggests that this proline-rich sequence motif might not be a precise indicator of the catalytic activity of YcaO-domain proteins.



4.2.3.3.ii. Role of hydrolase enzyme AmiC

The gene deletion experiments showed that deletion of AmiC abolishes production of streptamidine, and no additional metabolites were observed that gave clues about the function of this protein. However, the hydrolase should hypothetically be responsible for removing the leader and follower peptide from the streptamidine core peptide. This should occur prior to amidine formation, as a free N-terminal amine on His1 is required for cyclisation with the Leu2 carbonyl (**Figure 65**).

4.2.3.3.iii. Role of the dehydrogenase enzyme AmiE

Although gene deletion experiments showed that the dehydrogenase is essential for streptamidine biosynthesis, it was unexpected that this enzyme should be required because of the lack of any oxidation in the metabolite. Due to overproduction of shunt metabolites in the $\Delta amiE$ mutant (**Figure 62**), it is possible that instead of a catalytic role, AmiE is fulfilling a key structural role for proper cyclisation activity. In order to assess whether AmiE is similar to other characterised azole-forming dehydrogenases, its sequence was compared with that of the dehydrogenase McbC, involved in biosynthesis of microcin B17 (**Figure 64**, analysis carried out by Andrew Truman). This revealed that the two proteins only share ~10% identity. However, Phyre2 (267) analysis indicated that AmiE has a HY motif that structurally aligns with the catalytic KY residues of McbC (360), indicating catalytic potential for AmiE (**Figure 68**). A possible explanation for the role of AmiE is that it is catalytic, but a reductase reverses this activity. Alternatively, AmiE could oxidise another part of the streptamidine precursor peptide that is hydrolysed from the streptamidine core region. In order to fully determine the role of this enzyme, further detailed biochemical experiments would be required such as *in vitro* characterisation of the protein.

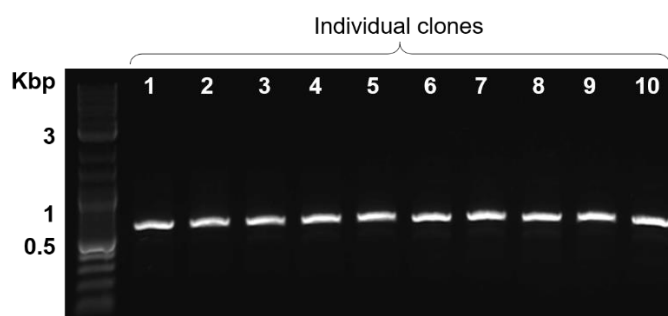


Figure 69: PCR screen of *E. coli* HME68-pCAPSalbC colonies transformed with a mutant oligonucleotide. Agarose gel image shows amplification of the target DNA region containing the precursor peptide gene from individual colonies.

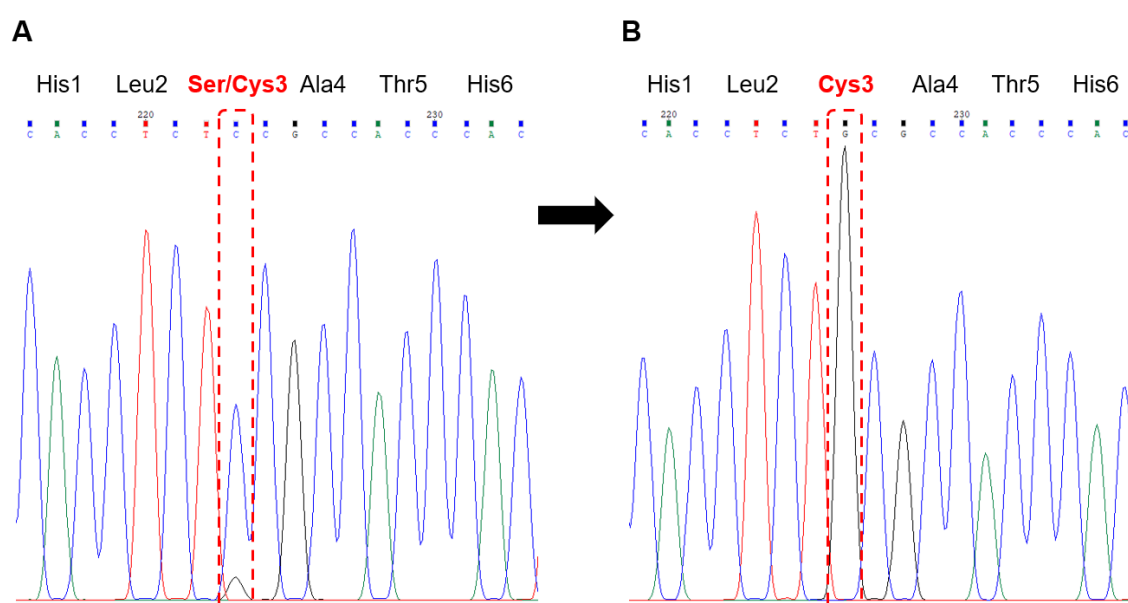


Figure 70: Sequence files obtained following PCR screen of colonies following mutagenesis of pCAPSalbC. A: Mixed plasmid population of “wild type” and mutant pCAPSalbC following transformation of *E. coli* HME68-pCAPSalbC with mutant oligonucleotide. B: Single population of clean pCAPSalbCSer3Cys mutant obtained following transformation of mixed plasmid population into *E. coli* DH5 α and screening of single colonies.

The mutated pCAPSalbC construct (pCAPSalbCSer3Cys) was transferred into *S. coelicolor* M1146 for metabolomic screening with comparison to *S. coelicolor* M1146-pCAPSalbC and *S. coelicolor* M1146-pCAPSalbC Δ amiA. However, due to time constraints the metabolomic profiles of these strains have not been fully analysed. The fermentation and analysis of these strains will therefore be carried out in the near future.

4.3.5. Prevalence of Streptamidines in Nature

As described in Chapter 2, homologues of the streptamidine biosynthetic gene cluster are widely distributed in nature, particularly in Actinobacteria. Some of the pathways contain identical biosynthetic machinery to streptamidine, and so will hypothetically produce an identical metabolite. Many of the gene clusters, however, harbour diverse genetic architectures and precursor peptide sequences. It is therefore likely that a suite of diverse, yet structurally related metabolites are produced by this novel gene cluster family. The gene clusters appeared to fall into two distinct subgroups, which correlated with the two sub-networks of precursor peptide sequences (**Figure 71**). After elucidating the structure of streptamidine, I could then compare the identified core peptide to that of similar precursor peptides (motif B). In total, I identified 11 variations of the core peptide sequence, from which I predicted the final structures of metabolites that might be produced by this gene cluster family (**Figure 72**). In all of the motif B-containing core peptides, the Ser3 residue is conserved, supporting the hypothesis that this residue plays an important role in biosynthesis.

The precursor peptides containing motif A correspond to biosynthetic gene clusters that contain two YcaO-domain proteins. It is therefore probable that the RiPP metabolites produced by these pathways harbour at least two separate YcaO-catalysed post-translational modifications. The fact that these precursor peptides are not rich in Ser, Thr or Cys residues is interesting, as these amino acids are usually transformed during YcaO-mediated catalysis such as azoline formation. It is therefore possible that one of these YcaO-domain proteins harbours unique catalytic potential. Many of the similar gene clusters also lack an encoded dehydrogenase, indicating that azole formation is unlikely. For this reason, it is difficult to predict what these structures might look like, and the pathways will require further characterisation.

Following precursor peptide-based predictions, I was also interested to see if streptamidine-like metabolites could be detected from public mass spectrometry datasets. To achieve this, MASST was utilised to carry out a search of mass spectral databases (analysis carried out by Andrew Truman). Interestingly, this analysis identified a metabolite with identical mass and MS² fragmentation to streptamidine in a marine actinomycete MS dataset collected by the Dorrestein lab (**Figure 73**), MassIVE MSV000078679). Although no further information could be found about the identity of this potentially related metabolite, this supports the idea that streptamidine-like gene clusters are widely distributed across actinomycetes in nature.

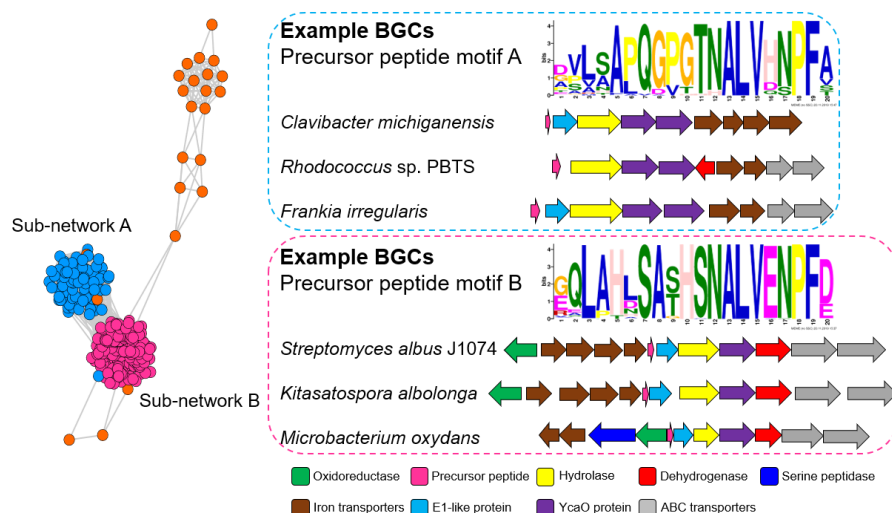


Figure 71: Networks of streptomidine-like precursor peptides identified from RiPPER analysis, in relation to the genetic architecture of corresponding biosynthetic gene clusters.

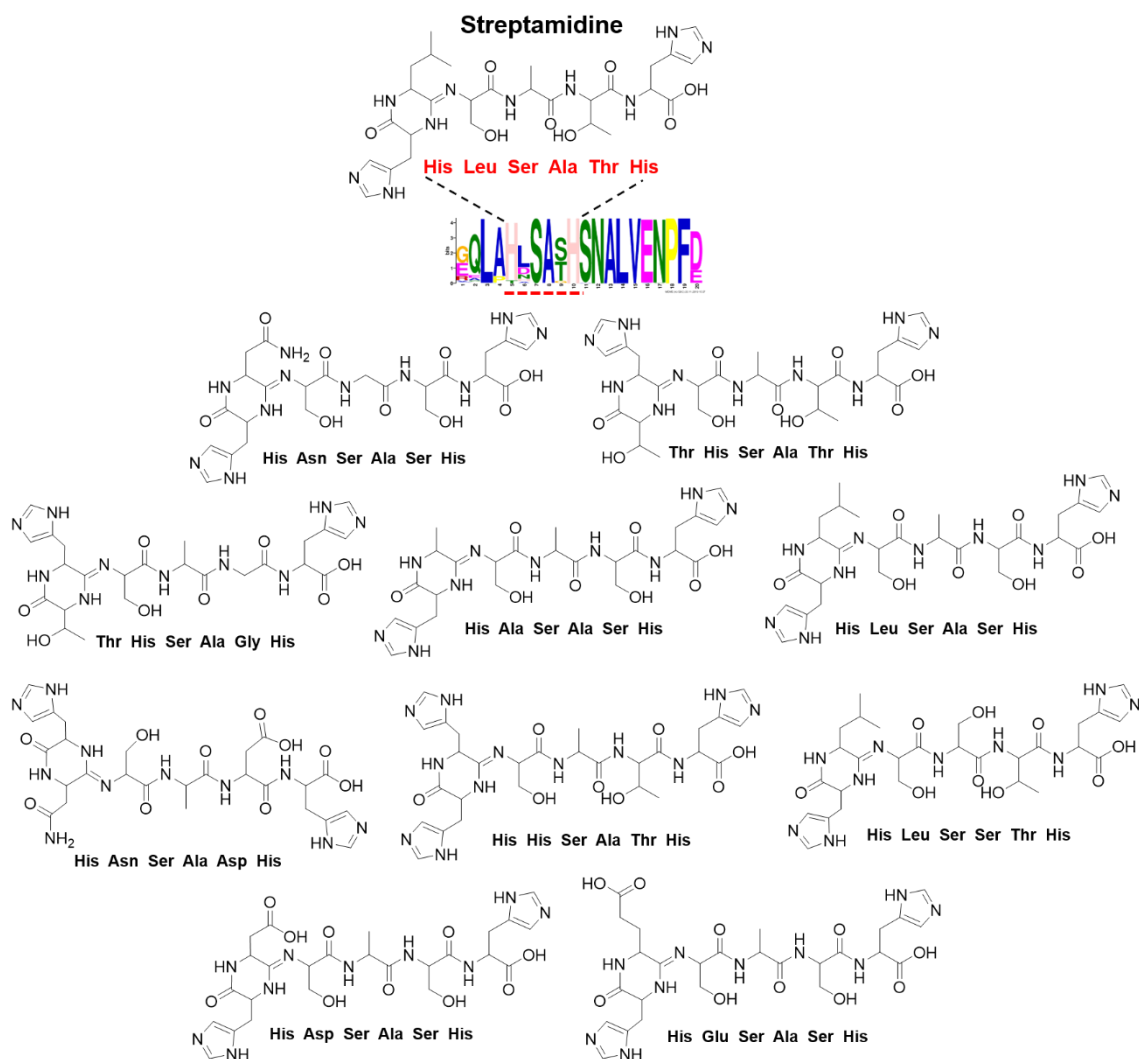


Figure 72: Putative structures of streptomidine-like metabolites based on variations of core peptide sequences. Inferred from precursor peptide sequences identified in homologous pathways.

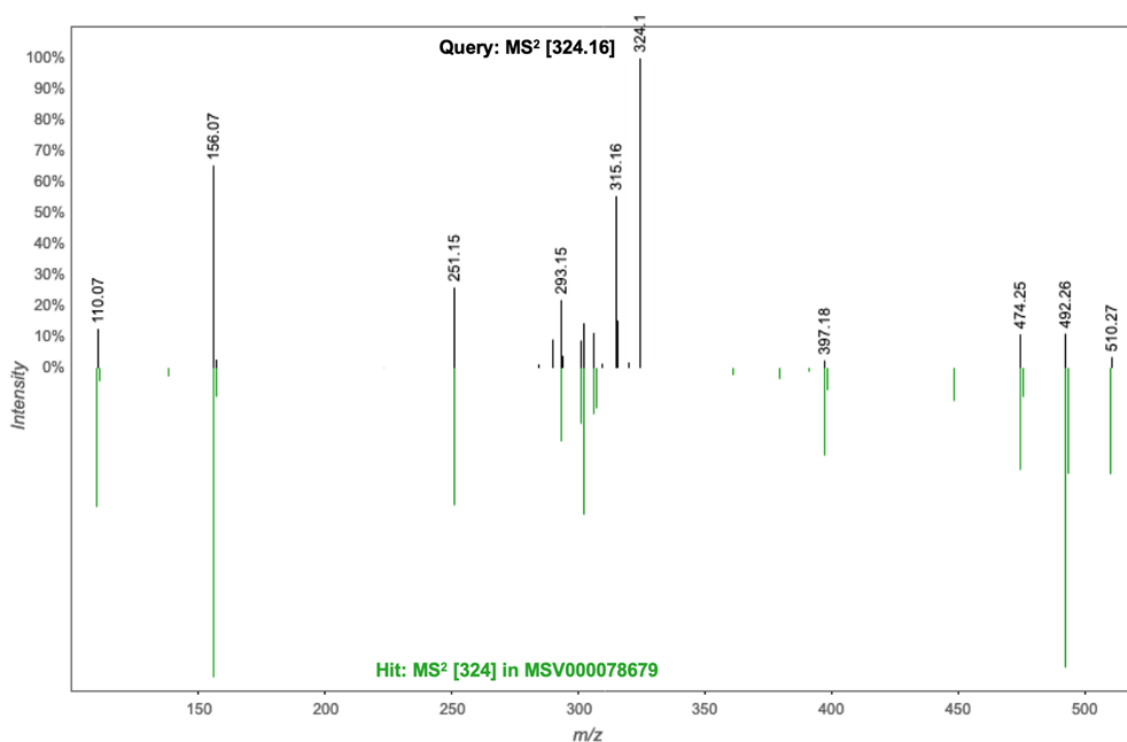


Figure 73: MASST analysis of streptomidine. Spectral match to the MS² spectrum of [streptomidine+2H]²⁺ (m/z 324.16) identified using MASST (Mass Spectrometry Search Tool) (361) at GNPS (Global Natural Products Social Molecular Networking). The hit is found in multiple samples of MassIVE Dataset MSV000078679 ("Zhang lab_microbes library_MS130001~9"), which is defined as an actinomycete dataset. The non-matching 324.1 relates to the unfragmented parent ion in the query spectrum.

4.3.6. Biological Role of Streptamidines in Nature

As streptamidine-related RiPPs are so widespread in nature, it is likely that they are playing an important biological or ecological role. Interestingly, klebsazolicin and bottromycin both act as antibiotics via ribosome inhibition. For both metabolites, the presence of the intact amidine ring has been shown to be essential for this activity, based on mutational studies of the biosynthetic machinery (129,204). This suggests that amidines might be an important structural feature for antimicrobial metabolites. With this in mind, I carried out a series of bioactivity assays with both pure streptamidine and with microbial extracts expressing the streptamidine gene cluster.

4.3.6.1. Antimicrobial assays

4.3.6.1.i. Assays with pure metabolite

Pure streptamidine was used for diffusion assays to test for microbial inhibition (method described in section 6.5.3.1.). Plugs were taken out of agar plates that had been overlaid with a range of indicator strains (**Table 3**), and wells were filled with 50 μ L streptamidine (1 mg/mL in H₂O) alongside 50 μ L of an appropriate positive control (kanamycin, apramycin or nalidixic acid (1 mg/mL)) and a negative control (H₂O). For all strains tested, no inhibition was observed, whilst inhibition was observed from the positive control antibiotics (**Figure 74**).

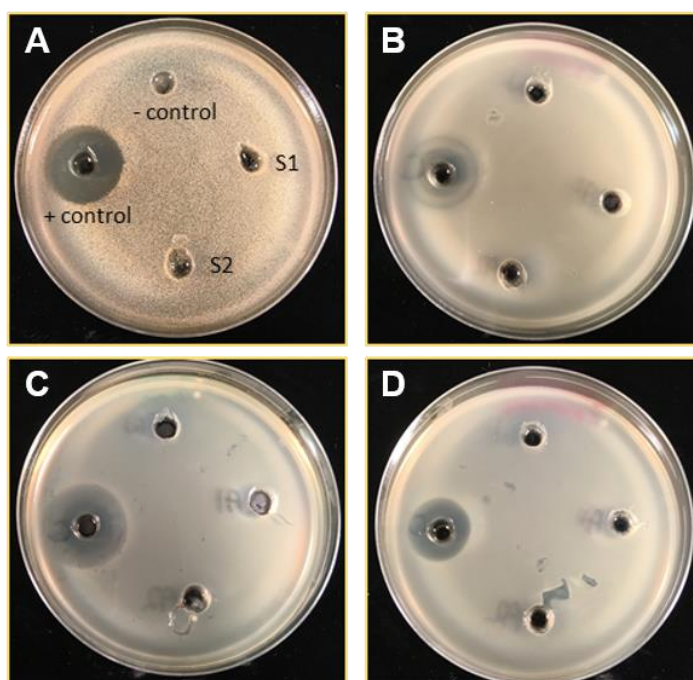


Figure 74: Bioassay plates generated during antimicrobial testing of streptamidine. Plugs were taken from agar plate overlaid with each indicator strain: *S. scabies* (A), *E. coli* ATCC25922 (B), *Pseudomonas fluorescens* (C) and *Bacillus subtilis* 168 (D), which were then separately loaded with 50 μ L streptamidine (1 mg/mL in H₂O), 50 μ L H₂O as a solvent control and 50 μ L of kanamycin (1 mg/mL) as a positive control. S1 and S2 represent two separate stocks of streptamidine from separate purifications.

4.3.6.1.ii. Competition assays

As well as testing pure streptomidine, I also carried out competition assays by spotting an inoculum of indicator strains alongside *S. albus* J1074, *S. albus*Δ*amiCD*, *S. coelicolor* M1146-pCAPS*albC* and *S. coelicolor* M1146-pCAPS*albC*Δ*amiA*. This was to confirm whether streptomidine, or any other metabolite produced by strains expressing the biosynthetic gene cluster, might display inhibitory activity. However, there was no observable difference in inhibition between positive and negative controls (**Figure 75**).

Table 3: List of microorganisms tested during bioactivity assays

Organism tested	Description	Inhibition observed
<i>Escherichia coli</i> ATCC25922	Gram-negative bacterium	None
<i>E. coli</i> NR986	Gram-negative bacterium (mutant with increased membrane permeability) (362)	None
<i>Pseudomonas aeruginosa</i> PA01	Gram-negative bacterium	None
<i>Pseudomonas fluorescens</i>	Gram-negative bacterium	None
<i>Bacillus subtilis</i> 168	Gram-positive bacterium	None
<i>Micrococcus luteus</i>	Gram-positive bacterium	None
<i>Mycobacterium smegmatis</i> MC2155	Gram-positive bacterium	None
<i>Streptomyces scabies</i>	Gram-positive bacterium	None
<i>Streptomyces cattleya</i>	Gram-positive bacterium	None
<i>Candida utilis</i>	Fungus	None

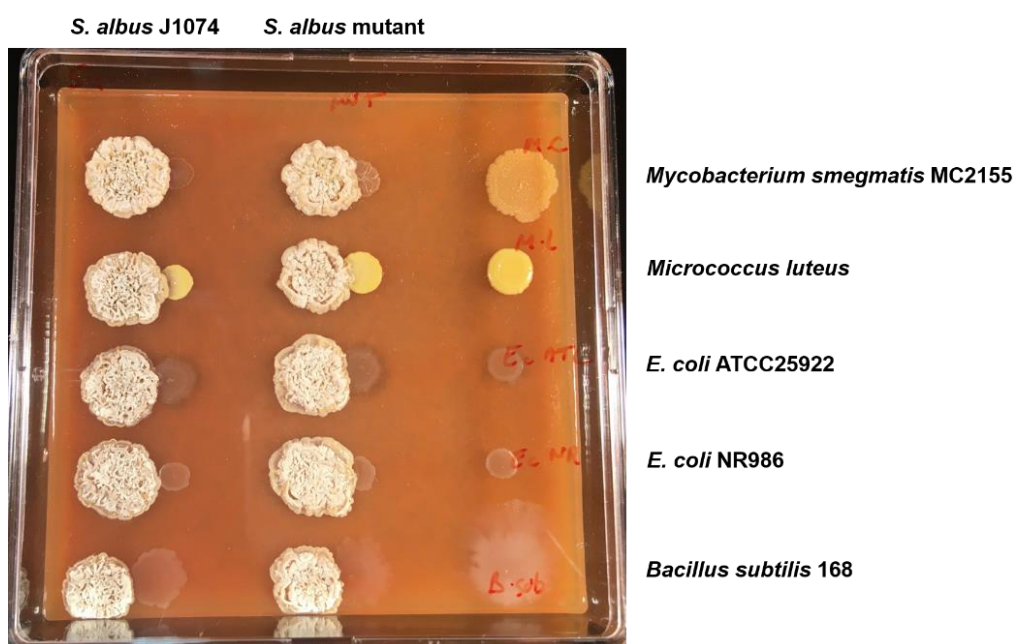


Figure 75: Bioassay plate generated during co-culture experiments. Extracts of *S. albus* J1074 and *S. albus* mutant were spotted on SM12 agar alongside extracts of a range of indicator strains to assess for antimicrobial activity of the streptomidine-producing strain.

4.3.6.2. Metal-binding assays

Although antimicrobial activity testing was inconclusive, another interesting feature of streptamidine biosynthesis is the presence of iron transporters in the biosynthetic pathway. Gene deletion experiments showed that these were essential for the production of streptamidine. Based on this, it was possible that streptamidine might exhibit metal-binding activity. To test for siderophore (iron-binding) activity, I first carried out a CAS assay (method described in section 6.5.3.2.). When mixing CAS solution with increasing concentrations of streptamidine, there was no observable colour change suggesting that streptamidine does not bind iron. To further test the hypothesis that streptamidine could bind metals, 20 μ L streptamidine (15 mM) was separately mixed with 500 μ L of a range of metal salts (FeCl_3 , CoCl_2 , CuCl_2 , MgCl_2 , MnSO_4 , NiSO_4 , ZnCl_2 ; 10 mM) and the solutions were analysed by LC-MS to see whether any metal-bound complexes were formed, based on the appearance of masses that correspond to a putative metal-bound complex. This approach has been successfully used to identify metal-bound complexes of other natural product siderophores in our lab (Javier Santos-Aberturas, unpublished). However, upon LC-MS analysis, no metal-bound complexes of streptamidine could be observed.

4.3.6.3. Developmental assays

Aside from antimicrobial and metal-binding activity, it is also possible that streptamidine plays a more subtle biological role in signalling and development. When growing *S. albus* J1074 wild type and mutant in various media, there was no phenotypic difference between the two strain. However, I was interested to see whether there was any phenotypic difference in growth under different stress-induced conditions. In order to test this, I made variations of SM12 agar medium containing high levels of salt (250 mM NaCl and 500 mM NaCl) to test growth under osmotic pressure (**Figure 76**), as well as variations of SM12 agar at a range of pHs (pH 4, pH 5, pH 6, pH 7 and pH 8) to test whether growth is affected under more acidic conditions. For each condition, I created serial dilutions of *S. albus* J1074 and *S. albus* Δ *amiCD* strains, which I spotted onto each agar test plate. After assessing growth over a period of 10 days, there did not appear to be any difference in growth between *S. albus* J1074 and *S. albus* Δ *amiCD* under any of these stress conditions.

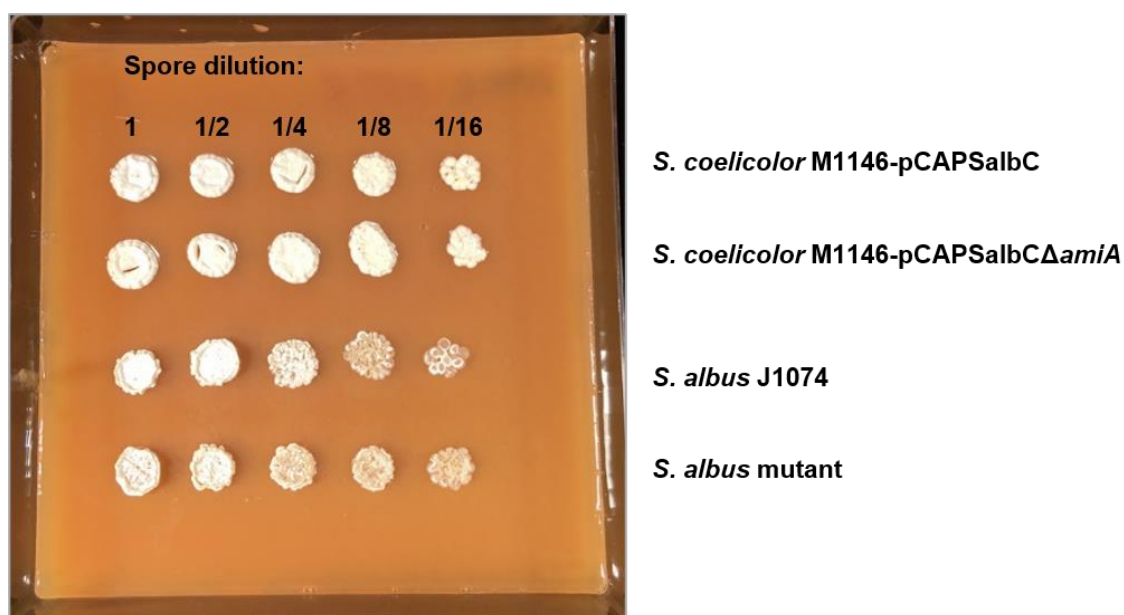


Figure 76: Spot assay plate generated during investigation of developmental effects of streptamidine. Serial dilutions of the two producing strains and two corresponding mutant strains were spotted onto SM12 agar containing a high salt concentration (500 mM NaCl) to assess whether streptamidine contributed any protective effect in comparison to the negative controls.

4.4. Chapter Summary

The work described in this chapter shows the characterisation of one member of a large related RiPP family. Streptamidine is a novel RiPP metabolite containing a structurally rare N-terminal amidine ring, which has previously only been characterised in the RiPPs klebsazolicin and bottromycin. Analysis of shunt metabolites and investigation of individual biosynthetic proteins within the streptamidine biosynthetic gene cluster allowed for predictions to be made about the roles of each of these RiPP tailoring enzymes in biosynthesis of streptamidine. Analysis of precursor peptide homologous with diverse core peptide sequences also highlighted that there are likely to be several analogues of streptamidine-like metabolites produced by diverse Actinobacteria. Although the bioactivity of this metabolite could not be determined through the experiments carried out, the widespread nature of this RiPP family suggest an important biological role for the producing species. Overall, this work suggests that amidine-containing RiPPs are much more prevalent than previously appreciated, and further investigation of this novel RiPP family will help shed light on the biosynthesis, structure and bioactivity of amidine-containing metabolites produced in nature.

Chapter 5: General Discussion and Future Work

5.1. Genome Mining for RiPPs

5.1.1. Streptamidine Discovery

The discovery of streptamidine was a result of utilising a tailored gene-led genome mining tool for RiPP natural products, searching for novel pathways containing the RiPP tailoring YcaO-domain enzymes. The success of this genome mining analysis in unveiling hundreds of previously uncharacterised biosynthetic gene clusters, including that of streptamidine which harbours a rare structural feature, demonstrates that bespoke RiPP genome mining is a promising approach for the discovery of structurally novel natural products in the future. It is particularly interesting that this gene cluster was discovered from strains such as *S. albus* J1074, whose metabolic capacity has been well studied (299,300) and yet streptamidine has not been previously reported. This suggests that genome mining has the potential to unveil many more biosynthetic gene clusters for natural products from both well-characterised species as well as from underexplored genera.

5.1.2. Future Work with RiPPER and YcaO-containing Pathways

Streptamidine is just one exemplar from a large family of novel RiPP metabolites. This thesis focused on the characterisation of this particular biosynthetic gene cluster in detail, but it would be interesting to investigate the products of related biosynthetic gene clusters in the future. It is particularly interesting that a group of homologous pathways contained two YcaO-domain proteins, suggesting that the resulting RiPPs contain additional post-translational modifications that might correspond to further structural novelty. Considering YcaO-domain proteins, the RiPPER analysis presented in Chapter 2 mapped a significant number of these enzymes to previously uncharacterised biosynthetic gene clusters, but the genetic context of many YcaO-domain proteins still remains uncharacterised. Therefore, further analysis of these enzymes could reveal even more biosynthetic and structural diversity. Attempts to study one of the identified biosynthetic gene clusters related to streptamidine from *Rhodococcus erythropolis* were carried out but early work was disrupted by the Covid-19 pandemic. As well as YcaO-domain proteins, RiPPER could also be used in the future to identify novel precursor peptides associated with a range of other RiPP tailoring enzymes used as the protein

'bait'. Based on the large amount of untapped diversity we identified from our analysis of Actinobacterial YcaO-domain proteins, it is highly likely that further analyses with RiPPER will uncover a vast number of other previously uncharacterised RiPPs. This in turn will help identify novel metabolites with potentially novel biochemical features.

5.2. Pathway Cloning and Engineering

5.2.1. Cloning and Genetic Manipulation

The streptamidine biosynthetic gene cluster was characterised after TAR cloning and heterologously expressing the pathway from *S. albus* J1074, and the native producer was also genetically manipulated to investigate pathway-related metabolism (described in Chapter 3). Although the streptamidine pathway was successfully cloned using TAR cloning, the process was challenging and yielded a high number of false-positive colonies despite the use of 5-FOA-based selection. The technique has also proved difficult for several other natural product gene clusters (personal communication from the lab group and other researchers at conferences). Further optimisation of this method may therefore be required to enable efficient cloning of gene clusters in the future, or other more recently developed methods could also be used, such as Cas9-assisted targeting of chromosome segments (CATCH) (327) or direct pathway cloning (DiPaC) (322). In the native producer of streptamidine (*S. albus* J1074), a homologous recombination-based disruption was carried out to produce a pathway mutant. Although this approach yielded complementary results to heterologous expression of the pathway, cleaner deletions of biosynthetic gene clusters could also be made using approaches such as CRISPR-Cas9-based genome editing. This particular approach is being increasingly improved for work in *Streptomyces* species (317,363). Individual gene deletions were successfully created within the TAR cloned streptamidine pathway using PCR-targeting (309), which is an efficient method for genetic deletions from plasmids. In order to study gene clusters in their native context, another alternative approach could be the use of CRISPR interference (CRISPRi) (364) to silence individual genes in the genomes of native producers of a gene cluster.

5.2.2. Synthetic Biology

RiPP gene clusters are highly amenable to synthetic biology approaches due to their biosynthetic logic. This has been demonstrated in this thesis for example, with the mutation of core peptide amino acids using mutant oligonucleotide-directed mutagenesis via the *mutS*-deficient strain *E. coli* HME68. Core peptide mutations could also be

introduced by expressing a mutant copy of the precursor gene in the pCAPSalbCΔ*amiA* mutant in an expression vector such as pIJ10257. However, as I was not able to successfully genetically complement this deletion with the native gene, I chose to utilise this alternative approach to mutagenesis. Previous work in our lab group has also demonstrated the use of yeast-mediated pathway refactoring to generate alternative metabolites related to a target RiPP such as bottromycin (331). Other RiPP families have also been subject to core peptide mutagenesis to generate analogues of the native metabolites. For example, Young *et al* utilised single-site mutagenesis for codon randomisation in thiopeptides (365) and Ruffner *et al* employed multiple site sequence randomisation to mutate the core peptide of the cyanobactin trunkamide (366). Similarly, a library of lanthipeptide analogues was created by Yang *et al* which helped to identify an inhibitor of protein-protein interactions (367). Together, genetic engineering and synthetic biology approaches can be seen as a chemical toolbox to help produce a range of ‘unnatural’ RiPP natural products. It was suggested by Travin *et al*, who discovered the antibiotic RiPP klebsazolicin, that it could be possible to rationally design hybrid antibiotics containing a common ribosome-targeting ‘warhead’, with variable uptake structures that control the specificity of the mode of action (129). Overall, there is exciting potential for the use of synthetic biology to engineer and alter RiPP pathways in order to produce alternative structures and improved bioactivity of natural products.

5.3. Structure and Biosynthesis of Streptamidine

5.3.1. Importance of Amidines in Nature

As shown in Chapter 4, streptamidine harbours a structurally rare amidine ring. As well as the RiPP antibiotics klebsazolicin and bottromycin (129,146), a handful of other non-RiPP amidine-containing metabolites with diverse bioactivities have also been reported from nature (**Figure 77**). Efraeptins are a group of peptides containing bicyclic amidines (368) that are produced by the soil hyphomycete fungus *Tolypocladium niveum* (369) via an NRPS pathway (370). Efraeptins display antifungal and insecticidal activity (371) and have also been shown to inhibit mitochondrial ATPase (372). A group of structurally similar metabolites called neofraeptins have also been reported, which are produced by the fungus *Geotrichum candidum* (368). These peptides also contain a bicyclic amidine at the C-terminus and display insecticidal activity (373). Other amidine-containing metabolites include ectoine, which contains a six-membered amidine (368) and is produced by several species of halophilic bacteria. L-aspartate-β-semialdehyde and L-glutamate are converted into ectoine via the intermediates 2-oxoglutarate and L-

2,4-diaminobutyrate (374). Ectoine acts as a protectant against osmotic and cold stress (375). Pyrostatins are metabolites produced by *Streptomyces* sp. SA-3501 that were originally reported to contain a five-membered amidine and inhibit N-acetyl-beta-D-glucosaminidase (376). This was recently refuted however, and a comparison of NMR data suggested that pyrostatins A and B are structurally identical to 5-hydroxectoine and ectoine respectively (377). Flustramine C is a marine natural product isolated from the bryozoan *Flustra foliacea*, which contains an amidine moiety within a brominated pyrroloindole structure (378). Several possible pathways have been suggested for the biogenesis of flustramine C via prenylation of the precursor deformylflustrabromine (379). Flustramine C is secreted from bryozoa into surrounding water and is thought to play an important role in ecological interactions (378), or to control bacterial growth on the surface of bryozoa (368). As well as the bioactivities of these metabolites, amidines may also play important roles in nature in the control of basicity, coordination and as a source of nitric oxide (368).

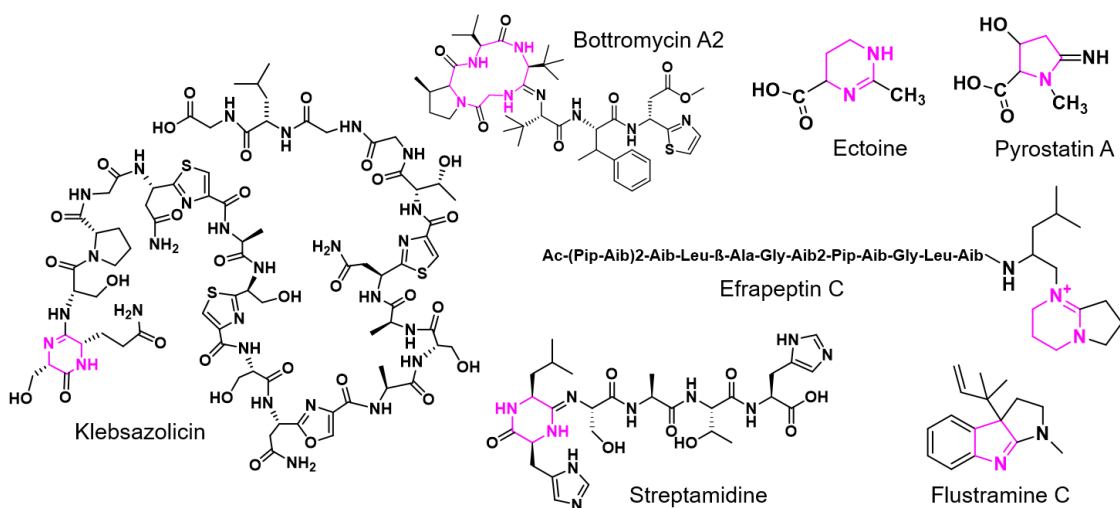


Figure 77: Chemical structures of amidine-containing metabolites: efrapeptin C, ectoine, pyrostatin A and flustramine C, alongside the RiPPs bottromycin, klebsazolicin and streptamidine.

5.3.2. Bioactivity of Streptamidine

One major limitation of genome mining for natural products is that there is no guarantee that an identified biosynthetic gene cluster will produce a metabolite with a particular bioactivity. This has been highlighted in this thesis, as the biological function of streptamidine was not determined from the bioactivity assays that were carried out. As discussed, characterised examples of amidine-containing metabolites are known to exhibit a range of important biological functions, and it is possible that streptamidine is playing an important biological role that was overlooked during this research. The bioassay experiments presented in this thesis do not represent an exhaustive list of

possible bioactivities, and this could therefore be further investigated in the future. For example, insecticidal, antiplasmodial and antiviral activities were not tested for, and there could be other developmental or signalling roles of streptamidine that are more difficult to determine under laboratory conditions. It is also possible that streptamidine displays narrow-spectrum antimicrobial activity, and that the range of indicator strains tested were not sufficient to detect this activity. The amidine-containing RiPPs bottromycin and klebsazolicin inhibit bacterial ribosomes, therefore it would be interesting to probe whether streptamidine has any specific ribosome-interfering activity. This could be tested using an *in vitro* reporter assay such as that used by Travin *et al* during investigation of klebsazolicin activity (129). A limitation of carrying out extensive bioactivity assays is that a large amount of pure metabolite is required. Streptamidine was difficult to purify in high quantities, therefore there was not a sufficient amount of pure metabolite to submit for wide-scale activity screening.

5.3.3. Biosynthesis of Streptamidine

In Chapter 4, a biosynthetic mechanism of streptamidine was proposed, based on results from gene deletion experiments and comparison of biosynthetic machinery in known RiPPs. As amidine-containing RiPPs are rare, further confirmation of streptamidine biosynthesis would be interesting to investigate. An avenue of research that was not fully explored during this thesis is the biochemical characterisation of proteins involved in the biosynthetic pathway. Expression and purification of individual proteins could give *in vitro* evidence that the precursor peptide and tailoring enzymes function together to produce streptamidine, and to determine the order of biosynthetic steps. Mutational analysis of key residues and analysis of interactions with the precursor peptide could also be carried out in the future. Unfortunately, a problem commonly occurred within the field is that YcaO-domain proteins are often insoluble and challenging to purify, making it difficult to study their biochemistry (personal communication within the lab group and other researchers). Further limitations of *in vitro* work are that reaction conditions might be too dissimilar to natural biosynthesis *in vivo*. For example, additional unknown co-factors might be required for certain enzymes to function, which could mean that the natural final product is not properly produced. Individually purified enzymes would also be present at a much higher concentration compared with natural conditions and might therefore behave in unusual ways. However, it would be interesting to complement the work presented in this thesis with further biochemical characterisation of the streptamidine biosynthetic pathway.

5.4. Conclusion

Overall, the work presented in this thesis demonstrates how a novel genome mining approach was utilised to unveil a large family of previously unknown RiPP metabolites from Actinobacteria, highlighting that we are still scratching the surface of the huge biosynthetic capabilities of microorganisms. The use of cloning and genetic tools led to the identification and characterisation of a structurally rare amidine-containing metabolite from *S. albus* J1074 and allowed biosynthetic proposals to be made. Future work investigating the biosynthesis and bioactivity of streptamidine, along with genome mining for further RiPP pathways, will help to increase our understanding about how these fascinating natural products are biosynthesised in nature, and why they are produced so widely in the microbial world. Furthermore, the use of amidine-containing metabolites as novel medicinal agents such as antibiotics represents an exciting potential for streptamidine and related RiPPs in the future, which is particularly important at a time when infectious disease is a pressing threat to public health.

Chapter 6: Materials and Methods

6.1. Materials

6.1.1. Strains

Table 4: Strains used during study.

Strain	Genotype/description	Application
<i>Saccharomyces cerevisiae</i> VL6-48N (380)	<i>MATα</i> , <i>his3-Δ1</i> , <i>trp1-Δ1</i> , <i>ura3-Δ1</i> , <i>lys2</i> , <i>ade2-101</i> , <i>met14 cir^o</i>	TAR cloning
<i>E. coli</i> DH5α (Invitrogen)	F ⁻ ϕ 80 <i>lacZ</i> Δ M15 Δ (<i>lacZ</i> YA- <i>argF</i>)U169 <i>recA1</i> <i>endA1</i> <i>hsdR17</i> (<i>r_k⁻</i> , <i>m_k⁺</i>) <i>phoA</i> supE44 <i>thi-1</i> <i>gyrA</i> 96 <i>relA1</i> λ ⁻	Transformation and maintenance of plasmids
<i>E. coli</i> ET12567/pUZ8002 (381)	<i>dam-13::Tn9 dcm-6 hsdM</i> Cml ^R , carrying helper plasmid pUZ8002	Conjugations involving Apra ^R constructs
<i>E. coli</i> ET12567/pR9604 (381)	<i>dam-13::Tn9 dcm-6 hsdM</i> Cml ^R , carrying helper plasmid pR9604	Conjugations involving Kan ^R constructs
<i>E. coli</i> DH5α/BT340 (309)	<i>E. coli</i> DH5 α carrying BT340 plasmid.	Gene deletions
<i>E. coli</i> BW25113/pIJ790 (309)	(Δ (<i>araD-araB</i>)567, Δ <i>lacZ</i> 4787(<i>::rrnB</i> - 4), <i>lacI</i> p-4000(<i>lacI</i> Q), λ -, <i>rpoS</i> 369(<i>Am</i>), <i>rph-1</i> , Δ (<i>rhaD-rhaB</i>)568, <i>hsdR</i> 514. Plasmid: pIJ790 [<i>oriR</i> 101], [<i>repA</i> 101(<i>ts</i>)], <i>araB</i> p-gam-be-exo	Gene deletions
<i>E. coli</i> HME68 (382)	W3110 <i>galK</i> tyr145UAG Δ <i>lacU</i> 169 [λ <i>cI</i> 857 Δ (<i>cro-bioA</i>)] Δ <i>mutS</i>	Mutagenesis of core peptide amino acids
<i>Streptomyces albus</i> J1074 (337)	Restriction-defective derivative (R ⁻ M ⁻) of <i>S. albus</i> G	Genetic source of streptamidine BGC
<i>Streptomyces coelicolor</i> M1146 (336)	Δ <i>act</i> Δ <i>red</i> Δ <i>cpk</i> Δ <i>cda</i>	Heterologous expression of gene cluster
<i>Streptomyces laurentii</i>	Wild type	Heterologous expression of gene cluster
<i>Streptomyces lividans</i>	Wild type	Heterologous expression of gene cluster

Table 5: *Streptomyces* strains constructed during study.

Strain	Description	Application
<i>Streptomyces albus</i> J1074ΔamiCD	<i>S. albus</i> J1074 pathway mutant with disrupted hydrolase and YcaO genes	Comparison with <i>S. albus</i> J1074 'wild type'
<i>Streptomyces lividans</i>-pCAPSalbC	<i>S. lividans</i> carrying cloned gene cluster from <i>S. albus</i> J1074	Heterologous expression of <i>S. albus</i> BGC
<i>Streptomyces laurentii</i>-pCAPSalbC	<i>S. laurentii</i> carrying cloned gene cluster from <i>S. albus</i> J1074	Heterologous expression of <i>S. albus</i> BGC
<i>Streptomyces coelicolor</i> M1146-pCAPSalbC	<i>S. coelicolor</i> M1146 carrying cloned gene cluster from <i>S. albus</i> J1074	Heterologous expression of <i>S. albus</i> BGC
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiA	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> BGC with precursor peptide gene deletion	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiB	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> BGC with E1-like gene deletion	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiC	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> BGC with hydrolase gene deletion	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiD	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> BGC with YcaO gene deletion	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiE	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> BGC with dehydrogenase gene deletion	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiX	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> BGC with oxidoreductase gene deletion	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiT1-2	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> BGC with combined ABC transporter gene deletions	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiF1-4	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> BGC with combined iron transporter gene deletions	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔMarR	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> BGC with MarR gene deletion	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔAT	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> J1074 BGC with combined acetyltransferase gene deletions	Gene deletion experiments

<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔPM	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> J1074 BGC with peptide methionine sulfoxide reductase gene deletion	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔOxy	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> J1074 BGC with oxygenase gene deletion	Gene deletion experiments
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiB-pIJamiB	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> J1074 BGC with complemented E1-like gene deletion	Complementations
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiB-pIJamiD	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> J1074 BGC with complemented YcaO gene deletion	Complementations
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCΔamiB-pIJamiE	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> J1074 BGC with complemented dehydrogenase gene deletion	Complementations
<i>Streptomyces coelicolor</i> M1146-pCAPSalbCSer3Cys	<i>S. coelicolor</i> M1146 carrying <i>S. albus</i> J1074 BGC with mutated core peptide (Ser3 to Cys3 codon change)	Core peptide mutagenesis

6.1.2. Plasmids

Table 6: Plasmids used during study.

Plasmid	Features	Resistance marker	Application
pCAP03 (332)	ARSH4/CEN6-Trp1, pUC ori, C31 int-attP-oriT-aph, URA3, ADH1	Kanamycin	TAR cloning
pKC1132 (383)	Conjugative vector, non-integrative, lacZa	Apramycin	<i>S. albus</i> pathway disruption
pIJ773ϕoriT (309)	oriT, non-conjugative, flippase recognition target (FRT) sites	Apramycin	Gene deletions via PCR targeting
pIJ10257 (384)	Φ BT1, <i>permE</i> *	Hygromycin	Genetic complementations

Table 7: Vectors constructed during study.

Construct	Resistance marker	Application
pSalbCAP	Kanamycin	Capture vector for TAR cloning of streptomidine BGC
pCAPSalbC	Kanamycin	Vector containing the TAR cloned streptomidine BGC
pCAPSalbCΔamiA	Kanamycin	pCAPSalbC with deletion of precursor peptide gene
pCAPSalbCΔamiD	Kanamycin	pCAPSalbC with deletion of YcaO gene
pCAPSalbCΔamiB	Kanamycin	pCAPSalbC with deletion of E1-like gene
pCAPSalbCΔamiE	Kanamycin	pCAPSalbC with deletion of dehydrogenase gene
pCAPSalbCΔamiC	Kanamycin	pCAPSalbC with deletion of hydrolase gene
pCAPSalbCΔamiX	Kanamycin	pCAPSalbC with deletion of oxidoreductase gene
pCAPSalbCΔamiT1-2	Kanamycin	pCAPSalbC with deletion of set of ABC transporter genes
pCAPSalbCΔamiF1-4	Kanamycin	pCAPSalbC with deletion of set of iron transporter genes
pCAPSalbCΔPM	Kanamycin	pCAPSalbC with deletion of peptidyl methionine gene
pCAPSalbCΔOxyg	Kanamycin	pCAPSalbC with deletion of oxygenase gene
pCAPSalbCΔAT	Kanamycin	pCAPSalbC with deletion of both acetyltransferase genes
pCAPSalbCSer3Cys	Kanamycin	pCAPSalbC with core peptide Ser3 to Cys3 mutation
pKCΔamiCD	Apramycin	Construct for disruption of <i>S. albus</i> BGC in native host
pIJamiD	Hygromycin	Construct for complementation of YcaO deletion
pIJamiB	Hygromycin	Construct for complementation of E1 deletion
pIJamiE	Hygromycin	Construct for complementation of dehydrogenase deletion

6.1.3. Chemicals and Media

Unless otherwise stated, all chemicals and media components were purchased from Sigma Aldrich, except for the following: agar (Melford), NaCl, glucose and sorbitol (Fisher Scientific), yeast extract (Merck), soya flour (Holland and Barrett) and peptone (BD Biosciences). Ultrapure water was obtained using a Milli-Q purification system (Merck). All media and was autoclaved prior to use and chemical solutions were filter sterilised using a 0.22 µm syringe filter.

Table 8: Antibiotics used during study.

Antibiotic	Solvent	Final concentration
Kanamycin	Water	50 µg/mL
Apramycin	Water	50 µg/mL
Carbenicillin	Water	50 µg/mL
Chloramphenicol	Ethanol	25 µg/mL
Nalidixic acid	Water	25 µg/mL
Hygromycin	PBS (phosphate-buffered saline)	50 µg/mL

Table 9: Solutions used for TAR cloning.

Solution	Ingredients per 100 mL, made up with milliQ H ₂ O
10x nitrogen bases	1.9 g YNB-AA, 1.9 g CSM-Trp, 5 g NH ₄ SO ₄
100x adenine	1 g adenine, 74 mM HCl
SPE	10 mM HEPES buffer pH 7.5, 100 mM EDTA pH 8, 18.2 g sorbitol
SOS	15 mM CaCl ₂ , 0.25 g yeast extract, 18.2 g sorbitol, 1 g peptone
STC	10 mM Tris/HCl pH 7.5, 10 mM CaCl ₂ , 18.2 g sorbitol
PEG	10 mM Tris/HCl pH 7.5, 10 mM CaCl ₂ , 20 g PEG8000

Table 10: Buffers used during study.

Buffer	Ingredients, made up with milliQ H ₂ O
SET buffer	75 mM NaCl, 20 mM Tris, 75 mM EDTA pH 7.2

Table 11: Media used during study.

Medium	Application	Ingredients per 1 L, made up with milliQ H ₂ O
LB (lysogeny broth)	<i>E. coli</i> , culture	10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl, adjusted to pH 7 with NaOH
L (lennox) broth	<i>E. coli</i> , culture	10 g tryptone, 5 g yeast extract, 5 g NaCl, 1 g glucose (John Innes Centre recipe)
DNA (difco nutrient agar)	<i>E. coli</i> , culture	4 g Difco Nutrient Broth powder, 10 g agar
SOB	<i>E. coli</i> , preparation of electrocompetent cells	20 g tryptone, 5 g yeast extract, 0.58 g NaCl, 0.186 g KCl, 2.03 g MgCl ₂ .6H ₂ O, 2.46 g MgSO ₄ .7H ₂ O
SOC	<i>E. coli</i> , chemical transformation of cells	20 g tryptone, 5 g yeast extract, 0.58 g NaCl, 0.186 g KCl, 2.03 g MgCl ₂ .6H ₂ O, 2.46 g MgSO ₄ .7H ₂ O, 3.6 g glucose
2xYT	<i>Streptomyces</i> , germination medium	16 g tryptone, 10 g yeast extract, 5 g NaCl, adjusted to pH 7.4 with NaOH
TSB (tryptic soy broth)	<i>Streptomyces</i> , culture	17 g tryptone, 3 g phytone, 5 g NaCl, 2.5 g K ₂ HPO ₄ , 2.5 g glucose
SFM (soya flour mannitol)	<i>Streptomyces</i> , sporulation medium	20 g mannitol, 20 g soy flour, 100 mM CaCl ₂ , 20 g agar
YPD (yeast peptone dextrose)	<i>Saccharomyces</i> , culture	10 g yeast extract, 20 g peptone, 20 g glucose (15 g agar), 0.004% adenine
SD-Trp (synthetic defined-tryptophan)	<i>Saccharomyces</i> , culture	5 g (NH ₄) ₂ SO ₄ , 1.7 g YNB-AA, 20 g glucose, 0.74 g CSM-Trp (20 g agar), 0.004% adenine
Top selective agar	<i>Saccharomyces</i> , TAR cloning selection	182 g sorbitol, 22 g dextrose, 30 g agar, 0.0002% 5-FOA, 0.004% adenine
Bottom selective agar	<i>Saccharomyces</i> , TAR cloning selection	182 g sorbitol, 22 g dextrose, 20 g agar, 0.0002% 5-FOA, 0.004% adenine
R5	<i>Streptomyces</i> , fermentation medium	103 g sucrose, 0.25 g K ₂ SO ₄ , 10.12 g MgCl ₂ .6H ₂ O, 10 g glucose, 0.1 g casamino acids, 2 mL trace element solution, 5 g yeast extract, 5.73 g TES buffer
SM12 (screening medium 12)	<i>Streptomyces</i> , fermentation medium	10 g soy flour, 50 g glucose, 4 g peptone, 4 g beef extract, 1 g yeast extract, 2.5 g NaCl, 5 g CaCO ₃ , adjust to pH 7.6 with KOH

SM14 (screening medium 14)	<i>Streptomyces</i> , fermentation medium	10 g glucose, 20 g soy peptone, 5 g meat extract, 5 g NaCl, 0.01 g ZnSO ₄ ·7H ₂ O, adjusted to pH 7.0 with KOH
BPM (bottromycin production medium)	<i>Streptomyces</i> , fermentation medium	15 g starch, 5 g yeast extract, 10 g soy flour, 5 g NaCl, 3 g CaCO ₃ , 25 µg/mL CoCl ₂
MacConkey agar	<i>E. coli</i> , mutagenesis red/white selection	17 g peptone, 3 g protease peptone, 10 g lactose monohydrate, 1.5 g bile salts, 5 g NaCl, 0.03 g neutral red, 0.001 g crystal violet, 13.5 g agar

6.1.4. Oligonucleotides

All primers were ordered from Eurofins Genomics or Integrated DNA Technologies (IDT) and purified by HPSF. Screening primers were ordered with a synthesis scale of 0.01 µmol and longer oligonucleotides for PCR targeting and capture vector construction were ordered with a synthesis scale of 0.05 µmol. Oligonucleotides for precursor peptide mutagenesis were ordered as 4 nmol ‘ultramers’ from IDT.

Table 12: Primers used during study. Restriction enzyme cut sites are underlined.

Primer name	Sequence (5'-3')	Application	Restriction site
SalbCap_Fw	GCTGCCGGGCGGCT <u>CCTAGG</u> TCTA CATCGGGGACATCAGCGACGCCCGT CCCGCGAGTCTTCCGATGCCGTTAAT <u>TAAGCC</u> ACTATTTATACCATGGGAGG CGTCAAC	Construction of pCAP03-derived capture vector for TAR cloning	AvrII, NdeI
SalbCap_Rv	TGTCCCCGATGTAGACCTAGGAGCC GGCCCGGCAGCTGACGGGTCAGCC ACGGCAGGAACCGCGGGCCGT <u>CATA</u> <u>TGTCG</u> AAAGCTACATATAAGGAACGT GCTG		AvrII, PacI
Salb_ClusScr_Fw	GCAGGACGGAACCGAGGGATG	Screening for cluster capture	
Salb_ClusScr_Rv	TGGGAGAGGATCGCCTCGGC		
Salb_YHMut_Fw	GATACAA <u>AGCTT</u> GACTGGATACGCG CCCAGC	Amplification of fragment for <i>S. albus</i> pathway disruption	HindIII
Salb_YHMut_Rv	GATACAGA <u>AATTC</u> GCTCACCTCCAGGC CGGACC		EcoRI
Salb_YHMut_ScrFw	GTGACCTGCGCGACCTCA	Screening <i>S. albus</i> disruption mutant	
Salb_YHMut_ScrRv	AGGGTGGCGTCAGGGGTCA		
Salb_TAR_PPDeI_Fw	CGTCCACCACGCATCGAACTGAATG GAGCTCAACT CATGATTCCGGGGATCCGTCGACC	Precursor peptide gene deletion	
Salb_TAR_PPDeI_Rv	TGTCAGCCGGCCGCCGTACCGGCG GCCTGGGCTGACTATGTAGGCTGGA GCTGCTTC		
Salb_TAR_E1DeI_Fw	GCCCATCCCCTCGTACTCCATCCGA CGGAGGTTTCCGTGATTCCGGGGAT CC	E1-like gene deletion	
Salb_TAR_E1DeI_Rv	CGGGCCGCTCCCCCTTGCGGGCAG CGGGGCCACCGCGGTGTAGGCTG GAGCTGCTTC		
Salb_TAR_YcaODeI_Fw	GACCTGCCGATGACCGCCGCCCTGC CCCTCGACGCCCTATTCCGGGGAT CCGTCGACC	YcaO gene deletion	

Salb__TAR_YcaODel_Rv	GGGAGCCGGGGGTACATGTGCGG GGCGGGGGCGAGGGTTGTAGGCTG GAGCTGCTTC		
Salb_TAR_HydrDel_Fw	GCTGCCCCGAAGGGGGAGCGGCC GGCATGACGCCCGCGATTCCGGGGA TCCGTCGACC	Hydrolase gene deletion	
Salb_TAR_HydrDel_Rv	GGGCGTCGAGGGGCAGGGCGGCGG TCATCGGCAGGTCTCTGTAGGCTGG AGCTGCTTC		
Salb_TAR_DehyDel_Fw	CGCCCCCGCCCCGCACATGTGACCC CCGGCTCCCCCATGATTCCGGGGAT C	Dehydrogenase gene deletion	
Salb_TAR_DehyDel_Rv	GGGTGAGGTGGTCGGGGGCGGGCC GTGCGCGGCGGCTCATGTAGGCTGG A		
Salb_TAR_OxidoDel_Fw	CGCCCGGCCCGCGACCCCTACCG AGGAGTTCCCGTGATTCCGGGGAT CCGTCGACC	Oxidoreductase gene deletion	
Salb_TAR_OxidoDel_Rv	GCGACACCCTGGCCCGCGCCTGGC CGAGCTGAGGAGTCATGTAGGCTGG AGCTGCTTC		
Salb_TAR_MarRDel_Fw	AGGGACGCTACACGACGAGCGAGGA GACCCGCGACCATGATTCCGGGGAT CCGTCGACC	MarR regulator gene deletion	
Salb_TAR_MarRDel_Rv	ACGGGGCGGAGGCGGACCCGGTGG GGCGGTGACTCCTCATGTAGGCTGG AGCTGCTTC		
Salb_TAR_Oxyg_Fw	TCGAAGTTCACCATCCAGCAGCGCG CGGTTCCCGCGATGATTCCGGGGAT CCGTCGACC	Oxygenase gene deletion	
Salb_TAR_Oxyg_Rv	CGTTCTCGCTCATGCGCGTCTCCTTC CTCGGCTCGTTCATGTAGGCTGGAG CTGCTTC		
Salb_TAR_IrTr_Fw	ACGCGTCCCCGAGGACGGAACCGA GGGATGAAGCCATGATTCCGGGGAT CCGTCGACC	Iron transporters gene deletions	
Salb_TAR_IrTr_Rv	GGTCGGTTCGTCGAGGAGGAGGGTC CGGGTGTCTGGGCTGTAGGCTGGA GCTGCTTC		
Salb_TAR_ABC_Fw	CGACCACCTACCCCGCCCGGAGAG ACGTACCCGCGATGATTCCGGGGAT CCGTCGACC	ABC transporters gene deletions	
Salb_TAR_ABC_Rv	CGTGTGCGTGCGCGTGTACGCACGC TTCGGTGCGGGTCATGTAGGCTGGA GCTGCTTC		
Salb_TAR_PepMet_Fw	CCGGGCGCATGTGATGCCAGTCGG GAGCACAGCGTATGATTCCGGGGAT CCGTCGACC	Peptide methionine sulfoxide reductase MsrA gene deletion	
Salb_TAR_PepMet_Rv	CCCCGGCTCCTCGGTCCGGTGAAGG AGTGCTGTGGCTCATGTAGGCTGGA GCTGCTTC		

Salb_TAR_Acet_Fw	CCGCCGGTCCGCACGCCACCGGGA GGGGCCCACCGCATGATTCCGGGA TCCGTCGACC	Acetyl transferase and maltose-O- acetyltransferase gene deletions	
Salb_TAR_Acet_Rv	GGGGCCGGCCCCCTTCGCGTGTACG TACGGGCCCCGTATGTAGGCTGGA GCTGCTTC		
SalbCycl_PE_Fw	GATACACATATGACCAGCAGCCGACT CGCC	Complementation of E1-like protein	NdeI
SalbCycl_PE_Rv	GATACAAAGCTTGTGGTCGCGGGCG TCATGC		HindIII
SalbYcaO_PE_Fw	GATACACATATGACCGCCGCCCTGC CC	Complementation of YcaO	NdeI
SalbYcaO_PE_Rv	GATACAAAGCTTGGGGAGCCGGGGG TCACATG		HindIII
SalbDehy_PE_Fw	GATACACATATGACCCCTGACGCCAC CCTCG	Complementation of dehydrogenase	NdeI
SalbDehy_PE_Rv	GATACAAAGCTTGCTCATCGGGCGG CTCCCAG		HindIII
SalbMarR_PE_Fw	GATACACATATGACGGCTGAGGACC GCCCGG	Complementation of MarR	NdeI
SalbMarR_PE_Rv	GATACAAAGCTTGTGACTCCTCAGCT CGGCCAGG		HindIII
pIJ10257_Fw	TTCGAGTGGCGGCTTGCG	Screening for pIJ10257 insert	
pIJ10257_Rv	CAAACGGCATTGAGCGTCAGC	Screening for pIJ10257 insert	
PPDel_screen_Fw	GCGGCTGGCCGGTCTGTTAC	Screening precursor peptide gene deletion	
PPDel_screen_Rv	CGGCTGCTGGTCACGGAAACC		
CyclDel_screen_Fw	GGTGCCGCGGACGACAAGTAG	Screening E1-like gene deletion	
CyclDel_screen_Rv	GTCGTACGGGTGCGGATCAG		
YcaODel_screen_Fw	ACCAGGCTGCGCGTCGAGA	Screening YcaO gene deletion	
YcaODel_screen_Rv	AGGTGGTCGAGGTCGACGGG		
HydrDel_screen_Fw	CTCCTACCGCCGACCTCCTC	Screening hydrolase gene deletion	
HydrDel_screen_Rv	GTGCCGAGCGAGACCCGGT		
DehyDel_screen_Fw	TCGACCTGACCACCGAGGACG	Screening dehydrogenase gene deletion	
DehyDel_screen_Rv	ACCAGGGCGAGCAGGGCG		
OxidoDel_Screen_Fw	CATCCCTCGGTTCCGTCTCG	Screening oxidoreductase gene deletion	
OxidoDel_Screen_Rv	AGCACCTGATCCGGCTGAC		
Salb_oxyg_scr_Fw	CAGTTGAGGGGCGGATCGTTC	Screening oxygenase deletion	
Salb_oxyg_scr_Rv	GAAAGGCCAGCTGGGCGTC		

Salb_IronTr_scr_Fw	CAGAGGCGTCCCACGCGTC	Screening iron transporters deletion	
Salb_IronTr_scr_Rv	GAACGGCGTGCGCGACTGCC		
Salb_ABCTr_scr_Fw	CTCACCCCGCCCGGAGAGAC	Screening ABC transporters deletion	
Salb_ABCTr_scr_Rv	GTGCGTGCGCGTGACGCAC		
Salb_PepMet_scr_Fw	GACCCCGGCTCCTCGGTCC	Screening peptide methionine sulfoxide reductase deletion	
Salb_PepMet_scr_Rv	CGCATGTCTGATGCCAGTCGG		
Salb_acet_scr_Fw	GTGACACCAAGGTGCCGCGAAC	Screening acetyl transferases deletions	
Salb_acet_scr_Rv	GGCCCCCTTCGCGTGACG		
MarR_screen_Fw	TCAGCCCGACCGGTCCTG	Screening MarR deletion	
MarR_screen_Rv	CGACCACGCCGAGGAGGTC		
SalbPP_Ser3CysRV	GTTCTCGACGAGGGCGTTGGAGTGG GTGGCGCAGAGGTGGGCCAGCTGG CCCGGGTCGGCGAT	Core peptide mutation	
Oligo100	AAGTCGCGGTCTGGAACCGTATTGCA GCAGCTTTATCATCTGCCGCTGGAC GGCGCACAAATCGCGCTTAA	Core peptide mutation	
SalbPPmut_FWScr	GCCGAAGTGCTGGGTGTCGAG	Screening core peptide mutation	
SalbPPmut_RVScr	GTGCCGGTGTCGAGGTGCAG		

6.2. General Methods

6.2.1. *E. coli*

6.2.1.1. Growth and maintenance

Unless otherwise specified, *E. coli* cells were grown in solid or liquid LB medium at 37 °C. 'Overnight' cultures and incubations were carried out for 16-18 hours or until colonies were visible on agar plates. Colonies on agar plates were stored at 4 °C and liquid culture stocks were stored in 50% glycerol at both -20 °C and -80 °C.

6.2.1.2. Transformation

Plasmid and single stranded DNA fragments were transformed into recipient *E. coli* cells via either chemical or electrical transformation, as stated in each protocol.

6.2.1.2.i. Chemical transformation

2-5 µL of DNA was used to transform 50-100 µL chemically competent cells. Cells mixed with DNA were incubated on ice for 2 minutes before being heat shocked for 45 seconds at 42 °C in a water bath. Cells were then immediately transferred back to ice for 2 minutes before adding 300 µL SOC medium. Cells were then incubated for 1 hour at 37 °C with shaking at 250 rpm. Between 100 µL-300 µL cell culture were then spread on LB agar plates with appropriate antibiotic selection.

6.2.1.2.ii. Making electrocompetent cells

Unless otherwise stated, electrocompetent cells were made as follows: a single colony or 30 µL of the desired cell stock was inoculated into 10 mL LB + 10 µL of the appropriate antibiotic and grown overnight at 37 °C with shaking at 250 rpm. 200 µL of the overnight culture was then used to inoculate 15 mL LB in a 50 mL Falcon™ tube + 15 µL of the appropriate antibiotic selection and grown at 37 °C until the OD_{600nm} reached between 0.2-0.4. Cells were harvested by centrifugation at 4,500 × *g* in a Sorvall® Biofuge primo centrifuge for 5 minutes at 4 °C and pellets were resuspended in 30 mL ice cold 10% glycerol. This step was repeated two more times, resuspending the pellets in 15 mL and 1 mL of 10% ice cold glycerol successively. 50 µL aliquots were transferred into microcentrifuge tubes and stored at -80 °C or used immediately for transformation.

6.2.1.2.iii. Electroporation

If previously prepared and stored at -80 °C, the vial of electrocompetent cells was thawed on ice for 3 minutes. 1-1.5 µL of DNA was added to 50 µL electrocompetent cells which were then transferred into a 2 mm electroporation cuvette. The outside of the cuvette

was dried, and it was inserted into a Bio-Rad Gene Pulser™ with the pulse generator set to 25 μ FD, 2.5 kV, and 200 Ω . The pulse was delivered, and the cuvette was immediately placed on ice before adding 1 mL cold LB. The cells were transferred back to a microcentrifuge tube and were incubated for an hour at 250 rpm, 37°C. Between 200 μ L-1 mL of the transformation mix was then spread on LB agar plates with appropriate antibiotic selection and grown overnight at 37°C.

6.2.1.3. Triparental mating

Individual 10 mL cultures of cells carrying a plasmid to be transformed and target recipient cells were grown overnight in the appropriate conditions. 20 μ L of each culture were then spotted together on an LB plate with no selection and grown at the appropriate temperature for four hours. The resulting patches of cells were picked with an inoculation loop and streaked onto an LB plate containing the appropriate antibiotic selection and incubated overnight at 37°C to obtain single colonies.

6.2.1.4. Plasmid isolation

Single colonies were picked and inoculated into 10 mL LB containing appropriate antibiotic selection and grown for 16 hours at 37 °C with shaking at 250 rpm. Cultures were then centrifuged for 5 minutes at 1,538 \times g in an Eppendorf Centrifuge 5810. Plasmid DNA was then extracted using a Promega Wizard Plus SV Minipreps DNA Purification System, following the corresponding protocol. DNA was eluted in 50 μ L H₂O.

6.2.2. *Streptomyces* Species

6.2.2.1. Growth and maintenance

Unless otherwise specified, all *Streptomyces* strains were grown at 28 °C in solid SFM for spore growth, solid SFM supplemented with 10 mM MgCl₂ for conjugations, liquid TSB for seed cultures and liquid SM12 media for fermentations. Other media used during screening trials include liquid R5, BPM and SM14 (**Table 11**). Liquid cultures were grown at 28 °C with shaking at 250 rpm. Spores and mycelium stocks were kept at -20 °C and -80 °C in 20% glycerol.

6.2.2.2. Genomic DNA extraction

For genomic DNA extraction from *Streptomyces* species, cultures were prepared by inoculating 40 μ L of spores from concentrated stocks in 50 mL TSB contained in 250 mL flasks with springs. Cultures were grown for 48 hours at 28 °C with shaking at 250 rpm. 30 mL of each culture was centrifuged for 5 minutes at 1,538 \times g in an Eppendorf Centrifuge 5810. Cell pellets were washed twice by resuspending in 5 mL H₂O and

centrifuging for 5 minutes at $1,538 \times g$ in an Eppendorf Centrifuge 5810. Mycelial pellets were stored at $-30\text{ }^{\circ}\text{C}$ overnight and then resuspended in 5 mL SET buffer (**Table 10**) with 100 μL lysozyme (50 mg/mL in H_2O) and incubated for 30 minutes at $37\text{ }^{\circ}\text{C}$. Following this, 140 μL proteinase K (20 mg/mL in H_2O) and 600 μL SDS (10% solution in H_2O) was added and mixed before incubating for 2 more hours at $55\text{ }^{\circ}\text{C}$ with occasional inversion. 2 mL of 5 mM NaCl was then added before cooling the suspension to $37\text{ }^{\circ}\text{C}$. 5 μL of chloroform was then added and the suspension was slowly mixed for 30 minutes at $20\text{ }^{\circ}\text{C}$. Cells were then centrifuged for 15 minutes at $4,500 \times g$ in a Sorvall® Biofuge primo centrifuge at $4\text{ }^{\circ}\text{C}$. The supernatant was slowly extracted into a fresh tube to which 2 mL H_2O and 5 mL chloroform was added. This was then centrifuged for 15 minutes at $4,500 \times g$ at $4\text{ }^{\circ}\text{C}$. The supernatant was transferred to a fresh tube to which 6 mL isopropanol was added and incubated for 5 minutes at room temperature. The DNA was then spooled with a sealed glass pipette and inoculated into 500 μL 70% ethanol. The ethanol was then extracted, and the DNA was left to air dry overnight before resuspending in 1 mL of H_2O .

6.2.2.3. Conjugation

Transfer of DNA from *E. coli* to *Streptomyces* cells was achieved via intergenic conjugation. Target DNA was first transformed into *E. coli* ET12567 cells containing a helper plasmid (pUZ8002 (KanR) or pR9604 (CarbR)). Single colonies were then picked to inoculate a 10 mL LB culture containing selective antibiotics for the donor strain, helper plasmid and incoming plasmid, which was grown overnight at $37\text{ }^{\circ}\text{C}$. 200 μL of the overnight culture was then used to inoculate a fresh LB culture containing the selective antibiotics, which was grown for approximately 4 hours at $37\text{ }^{\circ}\text{C}$. The *E. coli* cultures were then centrifuged for 5 minutes at $1,538 \times g$ in an Eppendorf Centrifuge 5810, the supernatant removed, and the cell pellets resuspended in 1 mL LB. Cells were then transferred to a 1.5 mL Eppendorf tube and centrifuged for 1 minute at $7,900 \times g$ in an Eppendorf Centrifuge 5424 R. This wash step was repeated once more, before resuspending the final cell pellet in 150 μL LB. Meanwhile, 30 μL of a concentrated stock of *Streptomyces* spores was mixed with 30 μL 2xYT medium and heat shocked for 10 minutes at $50\text{ }^{\circ}\text{C}$ in a water bath. Once cooled, the spores were mixed with the *E. coli* cells and 150 μL of the suspension was spread onto SFM agar plates containing 10 mM MgCl_2 . These plates were incubated at $28\text{ }^{\circ}\text{C}$ for 16-18 hours and each overlaid with 1 mL sterile H_2O containing 33 μL nalidixic acid and 33 μL of the selective antibiotic. Once dried, the plates were incubated at $28\text{ }^{\circ}\text{C}$ for 3-4 days or until exconjugants were visible. Single colonies were then picked and re-streaked onto SFM plates containing the selective antibiotic and incubated at $28\text{ }^{\circ}\text{C}$ until grown.

6.2.3. Cloning and Sequencing

6.2.3.1. Polymerase Chain Reaction (PCR)

6.2.3.1.i Analytical PCR from *Streptomyces exconjugants*

A patch of *Streptomyces* cells was picked with a sterile toothpick and inoculated in 180 μL 50 mM NaOH. Samples were then boiled for 10 minutes and neutralised with 20 μL 1 M Tris/HCl pH 8. The samples were then centrifuged for 15 seconds at $15,871 \times g$. The supernatant was then used as DNA template in a colony PCR detailed in **Tables 13** and **14**, using a Takara Terra PCR Direct Polymerase Mix.

Table 13: Reaction conditions for *Streptomyces* colony PCR with Takara Terra PCR Direct Polymerase mix.

Solution	Volume
Buffer	12.5 μL
Primer 1	0.75 μL
Primer 2	0.75 μL
Enzyme mix	0.5 μL
DNA template	2.5 μL
H ₂ O	8 μL
Total	25 μL

Table 14: PCR cycling conditions for *Streptomyces* colony PCR.

Step	Temperature	Time	Number of cycles
1	98 °C	2 min	x1
2	98 °C	10 sec	x34
3	60 °C	15 sec	
4	68 °C	2 min	
5	68 °C	3 min	x1
6	12 °C	Infinite hold	

6.2.3.1.ii. *E. coli* colony PCR

Single colonies from agar plates were picked with sterile toothpicks and inoculated in 50 μL sterile H_2O to be used as template DNA in a PCR. The PCR conditions shown in **Tables 15** and **16** were used with a Bio-Rad T100 ThermoCycler.

Table 15: Reaction conditions for *E. coli* colony PCR with G2 Taq polymerase.

Solution	Volume
GoTaq green buffer	5 μL
MgCl_2	2 μL
dNTPs 10 mM	0.5 μL
Primer 1 (forward) 10 pmol/ μL	0.5 μL
Primer 2 (reverse) 10 pmol/ μL	0.5 μL
G2 Taq DNA polymerase	0.25 μL
dH ₂ O	15.25 μL
Template DNA	1 μL
Total	25 μL

Table 16: PCR cycling conditions for *E. coli* colony PCR.

Step	Temperature	Time	Number of cycles
1	98 °C	3 min	x1
2	98 °C	20 sec	x30
3	60 °C	1 min	
4	72 °C	30 sec per kbp DNA	
5	72 °C	5 min	x1
6	4 °C	Infinite hold	

6.2.3.1.iii. High fidelity PCR for cloning

High fidelity cloning was carried out using Q5 DNA polymerase, with conditions as shown in **Table 17** and **Table 18** with a Bio-Rad T100 ThermoCycler.

Table 17: Reaction conditions for PCR cloning of genes from *Streptomyces* genomic DNA.

Solution	Volume
Q5 buffer	10 μ L
dNTPs	1 μ L
Template gDNA	1.5 μ L
Primer 1	2.5 μ L
Primer 2	2.5 μ L
GC enhancer	10 μ L
Q5 polymerase	0.5 μ L
H ₂ O	22 μ L
Total	50 μL

Table 18: PCR cycling conditions for cloning of genes from *Streptomyces* genomic DNA.

Step	Temperature	Time	Number of cycles
1	94 °C	2 min	x1
2	94 °C	45 sec	x35
3	55 °C	45 sec	
4	72 °C	20 sec per kbp DNA	
5	72 °C	5 min	x1
6	4 °C	Infinite hold	

6.2.3.2. Agarose gel electrophoresis

DNA samples were run on 0.8% agarose gels in 1x TBE (Tris/Borate/EDTA) buffer stained with 3 µL ethidium bromide per 100 mL agarose. DNA samples were loaded onto gels with 1 µL loading dye per 6 µL of sample, with reference to a 2-log 1 kbp DNA ladder (NEB). Samples were analysed under UV light.

6.2.3.3. Purification of DNA from agarose gel

DNA bands were excised from agarose gels and purified using a GE Healthcare illustra GFX PCR DNA and Gel Band Purification Kit following the supplied protocol, eluting the DNA in 20-30 µL elution buffer.

6.2.3.4. Purification of DNA from PCR mixtures

DNA amplified by PCR was purified directly from reaction mixtures using a Qiagen QIAquick PCR purification kit following the supplied protocol, eluting the DNA in 50 µL of elution buffer per 50 µL of initial PCR mixture.

6.2.3.5. DNA digestions

All restriction enzymes were purchased from New England Biolabs (NEB). DNA was digested with the appropriate restriction enzymes using 0.5 µL enzyme per 1 µg DNA. Digestions were made up to a 50-100 µL total volume of 10% CutSmart® Buffer in H₂O and incubated for 4 hours or overnight at 37 °C.

6.2.3.6. Ligations

An insert:vector ratio of 3:1 and 5:1 was calculated for each ligation reaction and the appropriate volume of digested insert and vector was mixed in a total reaction volume of 10 µL H₂O with 1 µL Invitrogen T4 Buffer and 1 µL Invitrogen T4 DNA ligase. Reactions were incubated at room temperature for 3 hours before being transformed into competent DH5α cells.

6.2.3.7. Sequencing

All DNA samples were sequenced by Eurofins genomics using a Mix2Seq Kit with the appropriate sequencing primers (**Table 12**). For plasmids, 15 µL of DNA at a concentration of 100 ng/µL was added to 2 µL of primer at a concentration of 10 pmol/µL. DNA purified from PCR mixtures were sequenced with the same primers that were used to amplify the DNA fragment in the PCR. In each case 15 µL of sample at a concentration of 5 ng/µL was added to 2 µL of primer at a concentration of 10 pmol/µL. DNA sequences

were analysed using NCBI nucleotide BLAST, Vector NTI and AlignX. Sequencing chromatograms were analysed using Chromas.

6.2.4. Mass Spectrometry

6.2.4.1. Preparation of mass spectrometry samples from culture extracts

Unless otherwise stated, samples from bacterial culture extracts for LC-MS analysis were prepared as follows: 1 mL of culture was mixed with one volume of methanol (MeOH) and agitated for 30 min at room temperature. The mixture was then centrifuged for 5.5 min at $15,871 \times g$ to remove debris, and 800 μ L of the supernatant was transferred to a 2 mL glass LC-MS vial for subsequent analysis.

6.2.4.2. Standard LC-MS analysis

Unless otherwise stated, all LC-MS analysis was carried out on a Shimadzu Nexera X2 UHPLC coupled to a Shimadzu IT-TOF mass spectrometer. Samples were injected onto a Phenomenex Luna Omega 1.6- μ m Polar C18 column (50 mm by 2.1 mm, 100 Å) set at a temperature of 40 °C and eluting with a linear gradient of (B) methanol from 0–60% in (A) H₂O + 0.1% formic acid over 6 minutes with a flow rate of 0.6 mL/min. Positive mode mass spectrometry data was collected between m/z 200 and 2,000. MS² data was also collected.

6.2.4.3. High resolution LC-MS² analysis

High resolution LC-MS² data were acquired by Gerhard Saalbach and Carlo de Oliveira Martins (John Innes Centre). Data were acquired on a Synapt G2-Si mass spectrometer equipped with an Acquity UPLC (Waters). Samples were injected onto an Acquity UPLC BEH C18 column, 1.7 μ m, 1 \times 100 mm (Waters) and eluted with a gradient of (B) acetonitrile/0.1% formic acid in (A) H₂O/0.1% formic acid with a flow rate of 0.08 mL/min at 45 °C. The concentration of B was kept at 1% for 1 min followed by a gradient up to 60% B over 10 min, and up to 99% over 2 min MS data were collected with the following parameters: resolution mode, positive ion mode, scan time 0.5 s, mass range m/z 50–1200 (calibrated with sodium formate), capillary voltage = 3.0 kV; cone voltage = 40 V; source temperature = 110 °C; desolvation temperature = 250 °C. Leu-enkephalin was used to generate a lock-mass calibration with m/z = 556.2766 measured every 30 seconds during the run.

6.3. Bioinformatics and Genome Mining

6.3.1. Analysis of YcaO-domain Proteins

6.3.1.1. Retrieval of YcaO-domain proteins from Actinobacteria

All standalone YcaO-domain proteins from Actinobacteria were identified in NCBI Genbank using CDART (Conserved Domain Architecture Retrieval Tool) (301). Sequences were retrieved and proteins smaller than 350 AA were excluded from further analysis. The remaining proteins were analysed and filtered using EFI-EST (302), selecting a sequence similarity network (SSN) with a 95% maximum identity cut-off. The SSN was visualised with Cytoscape software version 2.8.3 (385). Accession numbers of all proteins in the SSN were submitted to Batch Entrez to retrieve sequence files for further analysis.

6.3.1.2. Sequence alignments and phylogenetic tree building

The YcaO protein sequence files were aligned using MUSCLE (264) on the CIPRES Science Gateway (<https://www.phylo.org/>). This alignment was then used to construct a maximum likelihood tree using RAxML-HPC2 on XSEDE with default settings on the CIPRES Science Gateway. The tree was visualised and annotated with the interactive Tree Of Life (iTOL) (386).

6.3.2. Genome Mining

6.3.2.1. RiPPER

The 1,514 YcaO-domain protein sequences that were obtained were used as the input for RiPPER (<https://github.com/streptomyces/ripper>) (124) with default settings. The captured genomic regions were visualised in Artemis (261).

6.3.2.2. Whole genome comparative analysis

antiSMASH v5, PRISM v4, RiPPMiner, BAGEL v4 and DeepRiPP (full workflow) were used to analyse the whole genome sequence of *Streptomyces albus* J1074, using default settings.

6.3.3. Analysis of Precursor Peptides

6.3.3.1. Precursor peptide analysis

RiPPMiner (peptide) and DeepRiPP (NLPPrecursor) were used to analyse precursor peptide sequences using default settings.

6.3.3.2. Sequence alignments

Multiple sequence alignments of precursor peptides were performed using Clustal W (265) via the MEGA7 software (387), with gaps manually removed. Default alignment settings were employed (multiple alignment: gap opening penalty= 10, gap extension penalty=0.2; protein weight matrix= Gonnet, residue-specific penalties= ON, hydrophobic penalties= ON, gap separation distance= 4, end gap separation= OFF, use negative matrix= OFF, delay divergent cutoff= 30%).

6.3.3.3. Identification of motifs

Sequence motifs were searched for using the MEME tool (388) in the MEME suite version 5.1.1 (<https://meme-suite.org/index.html>). The settings were employed as follows: classic mode, site distribution= 'Any Number of Repetitions' number of motifs to search for= 5 or 1.

6.3.3.4. Evolutionary networking

Peptide similarity networking of the precursor peptide sequences were created using EGN (Evolutionary Gene and genome Network) (298) and visualised with Cytoscape version 2.8.3 (385).

6.3.3.5. Comparison of gene cluster architectures

The *S. albus* J1074 putative minimal biosynthetic gene cluster was submitted to MultiGeneBlast (<https://multigeneblast.sourceforge.net/>) (308) to identify homologous gene clusters.

6.3.4. Analysis of *S. albus* J1074 Biosynthetic Gene Cluster

Biosynthetic protein functions were analysed using a combination of blastP (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) (266) and Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) (389).

6.4. Characterising the *S. albus* J1074 Gene Cluster

6.4.1. TAR Cloning

6.4.1.1. Construction of capture vector

A vector to capture the gene cluster from *S. albus* J1074 genomic DNA (gDNA) was constructed by Gibson assembly between a linearised pCAP03 vector (332) and two single-strand oligonucleotides (Salb_TAR_fw and Salb_TAR_rv) (**Table 12**). The forward and reverse oligonucleotides had 34 and 36 nucleotide homology sequences with pCAP03 respectively, and were designed to generate a vector with 50 and 49 nucleotide homology sequences with upstream and downstream regions of the gene cluster respectively, either side of an AvrII restriction site. pCAP03 was digested with XhoI and NdeI, and 100 ng linearised plasmid and 10 pmol of each oligonucleotide were incubated with 5 µL ligase-free Gibson assembly reaction mixture (100mM Tris/HCl pH 7.5, 10 mM MgCl₂, 0.2 mM each dNTPs, 10 mM DTT, 1 mM NAD, 5% PEG-8000, 0.1125 units T5 exonuclease, 0.375 units Phusion polymerase, 10 µL total reaction volume) (**Table 19**) and incubated at 50 °C for 2 hours in a Bio-Rad T100 ThermoCycler. 10 µL of each assembly reaction was then transformed into *E. coli* DH5α by chemical transformation. Transformants were selected on LB + Kan agar.

Table 19: Gibson assembly reaction mixtures.

Reaction	Digested pCAP03	Primer 1 Salb_TAR_fw	Primer 2 Salb_TAR_rv	G.A. mastermix	H ₂ O	Total
1	3 µL (100 ng)	1 µL (0.1 pmol)	1 µL (0.1 pmol)	5 µL	-	10 µL
2	3 µL (100 ng)	1 µL (1 pmol)	1 µL (1 pmol)	5 µL	-	10 µL
3	3 µL (100 ng)	1 µL (10 pmol)	1 µL (10 pmol)	5 µL	-	10 µL
Control	3 µL (100 ng)	-	-	5 µL	2 µL	10 µL

Colonies containing the correctly assembled capture vector were identified by analytical PCR using the screening primers pCAP_sp and pCAP_asp. Plasmid DNA was isolated and sequenced using the same primers.

6.4.1.2. DNA digestion

Genomic DNA from *S. albus* J1074 was digested with NsiI and SmaI, and the capture vector was linearised between the capture arms with AvrII. DNA was digested with the appropriate restriction enzymes for 5 hours at 37 °C with occasional inversion according to **Table 20**.

Table 20: DNA digestion for TAR cloning.

Digestion	DNA amount	AvrII	NsiI	SmlI	CutSmart buffer	H ₂ O
<i>S. albus</i> gDNA	25 µg		15 µL	15 µL	50 µL	Up to 500 µL
pSalbCap	25 µg	15 µL			50 µL	Up to 500 µL

6.4.1.3. Spheroplast transformation

Digested material was transformed into *S. cerevisiae* VL6–48N by spheroplast polyethylene glycol (PEG)-8000 transformation, using the following method adapted from the Moore TAR 2.0 protocol (335):

Fresh yeast colonies were obtained by streaking *S. cerevisiae* onto YPD agar plates and incubating at 30 °C for 3 days. A yeast starter culture was made by inoculating a single colony into 10 mL YPD liquid medium and incubating at 28 °C for 16 hours with shaking at 250 rpm. 1 mL of starter culture was then inoculated in 50 mL YPD in a 250 mL flask and incubated at 28 °C for 6 hours with shaking at 250 rpm, at which point the OD_{600nm} had reached 0.85. The culture was then incubated on ice for 10 minutes before being transferred to a 50 mL Falcon™ tube and centrifuging for 3 minutes at 1,800 × g, 4 °C. The cell pellet was resuspended in 50 mL ice cold H₂O and centrifuged for 3 minutes at 1800 × g, 4 °C. The cell pellet was then resuspended in 50 mL ice-cold 1 M sorbitol and incubated at 4 °C overnight. The cell suspension was then inverted and centrifuged for 3 minutes at 1,800 × g, 4 °C. The cell pellet was resuspended in 20 mL SPE (**Table 9**), 40 µL 2-mercaptoethanol (BME) and 5 µL zymolyase (5 U/µL, Zymo Research). Cells were then incubated for 30 minutes at 30 °C and inverted every 5 minutes to reach a spheroplast level of 80-95%. The level of spheroplasting was quantified by comparing the OD_{600nm} of cells mixed with 1 M sorbitol and 2% SDS in 1:5 ratios. Cold 1 M sorbitol was added up to a total volume of 50 mL, and cells were centrifuged for 10 minutes at 600 × g, 4 °C. The cell pellet was resuspended in 3 mL STC (**Table 9**) by slowly pipetting with a 10 mL pipette, and then incubated at room temperature for 10 minutes. 200 µL spheroplasts were mixed with 500 ng linearised capture vector and 3 µg digested gDNA per transformation and incubated at room temperature for 10 minutes. 800 µL 20% PEG-8000 was added to each transformation and inverted 10 times before being incubated at room temperature for 20 minutes. Cells were then microcentrifuged for 10 minutes at 700 × g, 4 °C. Cell pellets were resuspended in 800 µL SOS (**Table 9**) and incubated for 40 minutes at 30 °C. Each cell suspension was added to 15 mL melted top selective agar and poured onto bottom selective agar plates that were incubated at 30 °C for 5 days.

6.4.1.4. Screening for captured gene cluster

100 individual colonies which grew on the selective agar plates were picked and re-streaked as small patches onto fresh SD-Trp plates and incubated for 2 days at 30 °C. After incubation, a small region from each patched colony was picked and inoculated in 400 µL 1 M sorbitol. 20 colonies were pooled into one screening sample. 1 µL zymolyase was added to each pool and samples were incubated at 28 °C with shaking at 250 rpm for 1 hour. Samples were then boiled for 10 minutes in a heat block and microcentrifuged (15 seconds, 1,000 × g) and 1 µL supernatant from each pool was then used as DNA template for an analytical PCR screen. Positive controls were set up containing *S. albus* J1074 gDNA (108 ng/µL) as a template, using the screening primers Salb_TARscr_Fw and Salb_TARscr_Rv (**Table 12**). Plasmid DNA from four positive clones was isolated and transformed into electrocompetent *E. coli* DH5α. An analytical digest of the purified construct was also carried out with the restriction enzymes HindIII-HF and SrfI to confirm the presence of the entire biosynthetic gene cluster.

6.4.2. Heterologous Expression

E. coli ET12567/pR9604 was transformed with pCAPSalbC by electroporation, and transformants were then used to transfer pCAPSalbC into the heterologous hosts *S. coelicolor* M1146, *S. lividans* and *S. laurentii* by intergeneric conjugation (described in section 6.2.2.3.). Nalidixic acid and kanamycin-resistant exconjugants containing integrated pCAPSalbC were verified by analytical PCR using the screening primers Salb_TARscr_Fw and Salb_TARscr_Rv (**Table 12**).

6.4.3. Gene Cluster Disruptions

A fragment of DNA corresponding to the overlapping *amiC* and *amiD* genes was cloned by high fidelity PCR with Q5 polymerase (section 6.2.3.1.iii. High fidelity PCR for cloning.) using the primers Salb_YHMut_Fw and Salb_YHMut_Rv (**Table 12**). The amplified DNA fragment was purified from agarose gel and digested with EcoRI and HindIII. The pKC1132 plasmid was also digested with EcoRI and HindIII and ligated with the digested DNA insert according to **Table 21**.

Table 21: Ligation trial for pKC1132-based gene disruption constructs.

Digested pKC1132	Digested insert	Vector: insert ratio	5x T4 ligase buffer	T4 DNA ligase	H ₂ O	Total
1.4 µL (50 ng)	6 µL	1:3	2 µL	0.25 µL	0.35 µL	10 µL
1.4 µL (50 ng)	10 µL	1:5	4 µL	0.5 µL	4.1 µL	20 µL
1.4 µL (50 ng)	-	<i>Control</i>	2 µL	0.25 µL	6.35 µL	10 µL

The resulting DNA construct was isolated and transferred into *S. albus* J1074 via intergenic conjugation with *E. coli* ET12567/pUZ8002 selected with apramycin and nalidixic acid. Exconjugants resistant to apramycin were validated by colony PCR (section 6.2.3.1.i .) using screening primers Salb_YHMut_ScrFw and Salb_YHMut_ScrRv to confirm that the YcaO and hydrolase genes had been disrupted. These primers bind within the disrupted region to confirm whether or not the target genetic region is intact.

6.4.4. Gene Deletions

Mutations of the *S. albus* J1074 biosynthetic gene cluster were carried out using PCR-targeting (309). First, the pCAPSalbC construct was transformed into *E. coli* BW25113 cells carrying the λRed plasmid pIJ790 and transformants were selected on LB + Chlor + Kan + Apra grown at 30 °C. Resistance cassettes were then amplified by high-fidelity PCR from pIJ773-OriT, a version of pIJ773 modified to have the OriT removed. All primers used for gene deletions are listed in **Table 12**. The forward primer was designed to contain a 20 bp sequence homologous to the start of the apramycin resistance gene (*aac(3)/IV*) alongside a 59 bp sequence homologous to the gene upstream of the target gene deletion, including the ATG start codon of the target gene. The reverse primer was designed to contain a 19 bp sequence homologous to the end of *aac(3)/IV* alongside a 58 bp sequence homologous to the gene downstream to the target gene deletion, including the ACT reverse stop codon of the target gene. Amplified resistance cassettes were purified and transformed into *E. coli* BW25113/pIJ790/pCAPSalbC and transformants were selected on LB + Chlor + Kan + Apra grown at 37 °C. Plasmid DNA was isolated from single colonies and transformed into *E. coli* DH5α cells carrying the BT340 plasmid. Flp-FRT recombination-mediated excision of the disruption cassette was then induced by streaking single colonies onto LB agar and incubating overnight at 42 °C. Single colonies in which the excision had taken place were screened for by patching colonies onto both LB+ Kan and LB+ Apra + Kan agar plates. The PCR-targeting mutant versions of pCAPSalbC were introduced into *S. coelicolor* M1146 by *E. coli* ET12567/pR9604-mediated intergeneric conjugation and selected by resistance to nalidixic acid and kanamycin.

6.4.5. Complementations

Constructs for the complementation of mutants were obtained by high-fidelity PCR amplification of each of these genes, using the primers listed in **Table 12**. The PCR products were digested with NdeI and HindIII and cloning by ligation into pIJ10257 (HygR) (384) digested with NdeI and HindIII. Ligation mixtures were transformed into chemically competent *E. coli* DH5 α and the plasmids were isolated and sequenced. The constructs were introduced into the corresponding *S. coelicolor* M1146-pCAPSalbC mutants by *E. coli* ET12567/pR9604-mediated intergeneric conjugation. Exconjugants were selected by resistance to nalidixic acid, kanamycin and hygromycin.

6.4.6. Core Peptide Mutations

Amino acids within the streptamidine core peptide were mutated using a Lambda-Red-mediated recombination strategy in *E. coli* cells deficient in mismatch repair. *E. coli* HME68 cells were grown in 10 mL LB with chloramphenicol and grown overnight at 30 °C with shaking at 250 rpm. 300 μ L of this culture was used to inoculate 15 mL LB with chloramphenicol, which was grown for four hours at 30 °C with shaking at 250 rpm. Recombineering activity was induced by incubating the cells for 20 minutes in a 42 °C water bath with shaking. Cells were then made electrocompetent and transformed with the pCAPSalbC construct. A 63 bp oligonucleotide (SalbPP_Ser3CysRV) was designed containing the desired amino acid mutation flanked by 30 bp regions of homology on either side. *E. coli* HME68-pCAPSalbC cells were then co-transformed with 1 μ L (4 pmol) of the mutant oligonucleotide and 1 μ L (4 pmol) of oligo100 (**Table 12**), which produces Gal⁺ recombinants of HME68 cells. Recombinant cells were selected for using red/white screening on MacConkey agar containing 1% galactose grown for 2 days at 30 °C. Red colonies were picked and re-streaked as patches on MacConkey agar containing 1% galactose and grown for a further 2 days at 30 °C. Mutations were then screened for by high-fidelity PCR using screening primers (SalbPPmut_FWScr and SalbPPmut_RVScr) binding ~300 bp each side of the desired mutation. DNA from PCR reactions was purified and sequenced, and mixed populations of 'wild type' and mutant strains were identified from sequencing data by identifying duplicated peaks for both the 'wild type' and mutant amino acid codon in the sequence chromatogram. Plasmid DNA from corresponding cells was isolated and transformed into *E. coli* DH5 α cells, and individual mutant clones were screened for as before. Mutant plasmids were transformed into corresponding *S. coelicolor* M1146 by *E. coli* ET12567/pR9604-mediated intergeneric conjugation.

6.4.7. Fermentation and Metabolomic Screening

6.4.7.1. Production cultures

Seed cultures of *S. coelicolor* M1146-pCAPSalbC were prepared by fermentation in a 50 mL flask containing 5 mL of TSB with kanamycin selection for 48 hours. 500 µL seed culture was used to inoculate 10 mL SM12, SM14, BPM and R5 in 50 mL Falcon tubes with caps replaced by foam bungs. Control strains carrying the TAR clone with a precursor peptide gene deletion were cultured in the same way for comparison. All fermentations were conducted in triplicate and incubated at 28 °C with shaking at 230 rpm. At day four of growth, samples were taken as described in section 6.2.4.2.

6.4.7.2. Untargeted metabolomic analysis

Untargeted comparative metabolomics was carried out on data from triplicate samples using Profiling Solution 1.1 (Shimadzu) with an ion m/z tolerance of 100 mDa, a retention time tolerance of 0.1 min and an ion intensity threshold of 70,000 units.

6.4.7.3. Metabolite networking

Global Natural Products Social Molecular Networking (100) (GNPS; <http://gnps.ucsd.edu>) was used to construct the metabolite networks from data acquired on the Shimadzu IT-TOF using the following GNPS settings: Parent Mass Tolerance = 1 Da, Min Pairs Cos = 0.6, Min Matched Peaks = 3, Network TopK = 15, MSCluster = ON, Minimum Peak Intensity = 25, Filter Precursor Window = OFF, Filter Library = OFF, and Filter peaks in 50 Da Window = OFF.

6.5. Purification and Characterisation of Streptamidine

6.5.1. Purification of Metabolite

6.5.1.1. Large scale cultures

Four 2-litre flasks containing 0.5 L of SM12 were each inoculated with 25 mL of *S. coelicolor* M1146-pCAPSalbC TSB seed culture grown for 48 hours at 28 °C. After four days fermentation at 28 °C with shaking at 250 rpm, the cultures were centrifuged to remove debris, combined and filtered to yield approximately 1.5 L of crude extract.

6.5.1.2. Liquid extraction

The crude extract was washed with ethyl acetate (3 x 1.5 litres) and the aqueous layer was separated from the organic layer. The aqueous layer was extracted with 1-butanol (3 x 1.0 litres) and the final aqueous layer was separated from the organic layer. The original solvent was removed using a Buchi rotary evaporator.

6.5.1.3. Solid phase extraction

The aqueous extract was concentrated *in vacuo* to approximately 50 mL and subjected to solid-phase extraction (SPE) chromatography on a 400 g SNAP Ultra HP20 cartridge connected to a Biotage using a gradient of H₂O-MeOH (100:0 to 0:100) at a flow rate of 20 mL/min. Fractions were analysed by LC-MS (as described in section 6.2.4.2.) and those containing streptamidine were combined, methanol was removed and the samples were freeze-dried.

6.5.1.4. High performance liquid chromatography (HPLC)

Samples obtained from Biotage purification were resuspended in H₂O and subject to semi-preparative HPLC using a Phenomenex Luna PFP(2) column, 5 µm, 250 x 10 mm, with a gradient of 0.1% aqueous formic acid-MeOH (98:2 to 90:10) over 35 minutes at a flow rate of 2 mL/min. Fractions were collected based on the appearance of peaks in the UV chromatogram monitored at 210 nm, and then analysed by LC-MS (as described in section 6.2.4.2.) to determine the retention time of streptamidine. Fractions containing streptamidine were combined and freeze dried. A final purification step was then carried out using a semi-preparative Luna Omega Polar C18 column, 5 µm, 250 x 10 mm, with an isocratic gradient of aqueous 0.1% formic acid-MeOH (90:10) for 16 minutes followed by a wash gradient from 90:10 to 5:95 over 5 minutes and a flow rate of 2.8 mL/min. Streptamidine eluted in two phases corresponding to two peaks in the UV chromatogram monitored at 210 nm, which were separately collected, analysed by LC-MS, freeze-dried and retained for NMR analysis.

6.5.2. Structural Elucidation

6.5.2.1. NMR (Nuclear Magnetic Resonance)

Pure streptomidine (1.4 mg) corresponding to the first HPLC UV peak was dissolved in 600 μL DMSO- d_6 from an individual vial and subjected to a series of 1D and 2D nuclear magnetic resonance (NMR) experiments on a Bruker Ascend 600 MHz instrument at 298 K. The NMR experiments carried out were Proton (64 scans), Carbon (25,000 scans), HSQCed (100 scans), HMBC (64 scans), COSY (16 scans), TOCSY (32 scans) and HSQC-TOCSY (64 scans). Spectra were analysed using Bruker TopSpin 3.5 and Mestrelab Research Mnova 14.0 software, with assistance from Sergey Nepogodiev.

6.5.2.2. Fluorescamine-binding assay

3 μL of fluorescamine ((4-phenylspiro-[furan-2(3H),1-phthalan]-3,3'-dione) (Sigma-Aldrich), 3 mg/mL in acetone) was separately mixed with 3 μL of L-histidine (1 mg/mL in H_2O) and 3 μL of streptomidine (1 mg/mL in H_2O), in a total volume of 500 μL H_2O within a 1.5 mL glass LC-MS vial. Samples were then analysed by LC-MS as described in section 6.2.4.2.

6.5.2.3. Marfey's analysis

500 μg of pure streptomidine was hydrolysed for 16 hours with 100 μL 6 M HCl at 100 $^\circ\text{C}$ in a sealed glass vial contained within a heated sand block. The HCl was dried off under a stream of N_2 for 1 hour at room temperature, with assistance from Martin Rejzek. The hydrolysed sample was mixed with 20 μL 1 M NaHCO_3 and 40 μL Marfey's reagent (L-FDAA ($\text{N}\alpha$ -(2,4-dinitro-5-fluorophenyl)-L-alaninamide) 1% solution in acetone) and incubated at 40 $^\circ\text{C}$ for 1 hour. The reaction was then neutralised with 20 μL 1 M HCl. The samples were diluted with 500 μL 50% acetonitrile and centrifuged for 1 minute at 15,871 $\times g$ to remove debris. For derivatisation of amino acid standards, 50 μL of each L- and D- amino acid (histidine, leucine, serine, alanine, threonine, 2 mg/mL in H_2O) was mixed with 20 μL 1 M NaHCO_3 and 40 μL L-FDAA in a 1.5 mL Eppendorf tube and incubated at 40 $^\circ\text{C}$ for 1 hour. The reaction was then neutralised with 20 μL 1 M HCl. The samples were diluted with 1 mL 50% acetonitrile and centrifuged for 1 minute at 15,871 $\times g$ to remove debris. 1 μL each sample were injected onto a Kinetix 1.6- μm C18 column (50 mm by 2.1 mm, 100 \AA) set at a temperature of 40 $^\circ\text{C}$ and eluting with a linear gradient of (B) 50% acetonitrile from 5-50% in (A) 0.1% formic acid in H_2O over 6 minutes with a flow-rate of 0.6 mL/min. Samples were analysed on a Shimadzu Nexera X2 UHPLC coupled to a Shimadzu IT-TOF mass spectrometer.

6.5.3. Bioassays

6.5.3.1. Antimicrobial assays

Ten mL cultures of each indicator strain (**Table 3**) were grown in LB (YPD for *C. utilis*) overnight at 37 °C (30 °C for *C. utilis*). 100 µL of each culture was then used to inoculate a 10 mL subculture of each strain in the same medium, which were grown for 5 hours at 37 °C (30 °C for *C. utilis*). 1 mL of each culture was then mixed with 14 mL molten LB agar, which was poured into plates. Once solidified, three 1 cm diameter plugs were taken from each agar plate, which were then separately loaded with 50 µL streptomidine (1 mg/mL in H₂O), 50 µL H₂O as a solvent control and 50 µL of an appropriate antibiotic/antifungal agent as a positive control (kanamycin, apramycin or nalidixic acid (1 mg/mL)). Plates were incubated overnight at 37 °C (30 °C for *C. utilis*).

6.5.3.2. Metal binding assays

6.5.3.2.i. CAS assay

For the CAS assay, 500 µL of CAS assay solution (prepared as described by Alexander and Zuberer (390)) was mixed with 10 µL increasing concentrations of streptomidine from 1.5 µM to 25 µM.

6.5.3.2.ii. LC-MS metal binding assays

For LC-MS binding assays, solutions of 10 mM metal salts were prepared (FeCl₃, CoCl₂, CuCl₂, MgCl₂, MnSO₄, NiSO₄, ZnCl₂, dissolved in 10 mM HCl) and 500 µL of each were mixed with 20 µL streptomidine (15 µM). Samples were then analysed by LC-MS as described in section 6.2.4.2.

6.5.4. MASST Analysis

Streptomidine-like metabolites were searched for using a MASST search (361) at GNPS (<https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>) with the following settings: parent mass tolerance = 2 Da; min matched peaks = 5; ion tolerance = 0.5 Da; score threshold = 0.7; library = specLib.

References

1. Emmerich R, Löw O. Bakteriolytische Enzyme als Ursache der erworbenen Immunität und die Heilung von Infektionskrankheiten durch dieselben. Zeitschrift für Hyg und Infect [Internet]. 1899;31(1):1–65. Available from: <https://doi.org/10.1007/BF02206499>
2. Rogers LA. the Inhibiting Effect of Streptococcus Lactis on Lactobacillus Bulgaricus. J Bacteriol [Internet]. 1928 Nov;16(5):321–5. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/16559344>
3. Dias DA, Urban S, Roessner U. A Historical overview of natural products in drug discovery. Metabolites [Internet]. 2012 Apr 16;2(2):303–36. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/24957513>
4. Manyi-Loh C, Mamphweli S, Meyer E, Okoh A. Antibiotic use in agriculture and its consequential resistance in environmental sources: Potential public health implications. Molecules [Internet]. 2018 Mar 30;23(4):795. Available from: <https://pubmed.ncbi.nlm.nih.gov/29601469>
5. Cantrell CL, Dayan FE, Duke SO. Natural products as sources for new pesticides. J Nat Prod [Internet]. 2012 Jun 22;75(6):1231–42. Available from: <https://doi.org/10.1021/np300024u>
6. Behie SW, Bonet B, Zacharia VM, McClung DJ, Traxler MF. Molecules to ecosystems: Actinomycete natural products in situ. Front Microbiol [Internet]. 2017;7(JAN):2149. Available from: <https://www.frontiersin.org/article/10.3389/fmicb.2016.02149>
7. Davies J. Specialized microbial metabolites: Functions and origins. J Antibiot (Tokyo) [Internet]. 2013;66(7):361–4. Available from: <https://doi.org/10.1038/ja.2013.61>
8. Currie CR, Scottt JA, Summerbell RC, Malloch D. Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. Nature [Internet]. 1999;398(6729):701–4. Available from: <https://doi.org/10.1038/19519>
9. Kroiss J, Kaltenpoth M, Schneider B, Schwinger MG, Hertweck C, Maddula RK, et al. Symbiotic streptomycetes provide antibiotic combination prophylaxis for wasp offspring. Nat Chem Biol. 2010 Apr;6(4):261–3.
10. Scott JJ, Oh DC, Yuceer MC, Klepzig KD, Clardy J, Currie CR. Bacterial protection of beetle-fungus mutualism. Science (80-). 2008 Oct;322(5898):63.
11. Bleich R, Watrous JD, Dorrestein PC, Bowers AA, Shank EA. Thiopeptide antibiotics stimulate biofilm formation in *Bacillus subtilis*. Proc Natl Acad Sci U S A [Internet]. 2015/02/23. 2015 Mar 10;112(10):3086–91. Available from:

- <https://pubmed.ncbi.nlm.nih.gov/25713360>
12. Grinberg M, Orevi T, Kashtan N. Bacterial surface colonization, preferential attachment and fitness under periodic stress. *PLoS Comput Biol* [Internet]. 2019 Mar 5;15(3):e1006815. Available from: <https://doi.org/10.1371/journal.pcbi.1006815>
 13. Fleming A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. 1929. *Bull World Health Organ* [Internet]. 2001;79(8):780–90. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/11545337>
 14. Chain E, Florey HW, Adelaide MB, Gardner AD, Oxford DM, Heatley NG, et al. Penicillin As a Chemotherapeutic Agent. *Lancet* [Internet]. 1940;236(6104):226–8. Available from: <http://www.sciencedirect.com/science/article/pii/S0140673601087281>
 15. Haas LF. Papyrus of Ebers and Smith. *J Neurol Neurosurg Psychiatry* [Internet]. 1999 Nov;67(5):578. Available from: <https://pubmed.ncbi.nlm.nih.gov/10519860>
 16. Dubern JF, Diggle SP. Quorum sensing by 2-alkyl-4-quinolones in *Pseudomonas aeruginosa* and other bacterial species. *Mol Biosyst*. 2008 Sep;4(9):882–8.
 17. BRUNEL J. Antibiosis from Pasteur to Fleming. *J Hist Med Allied Sci* [Internet]. 1951 Jul 1;6(3):287–301. Available from: <https://doi.org/10.1093/jhmas/VI.Summer.287>
 18. Gharsallaoui A, Oulahal N, Joly C, Degraeve P. Nisin as a Food Preservative: Part 1: Physicochemical Properties, Antimicrobial Activity, and Main Uses. *Crit Rev Food Sci Nutr*. 2016 Jun;56(8):1262–74.
 19. Dubos RJ. Studies on a bactericidal agent extracted from a soil bacillus: II. Protective effect of the bactericidal agent against experimental pneumococcus infections in mice. *J Exp Med*. 1939 Jun;70(1):11–8.
 20. Hotchkiss RD, Dubos RJ. Bactericidal Fractions From an Aerobic Sporulating Bacillus. *J Biol Chem* [Internet]. 1940 Dec 1;136(3):803–4. Available from: <http://www.jbc.org/content/136/3/803.short>
 21. Kresge N, Simoni RD, Hill RL. Selman Waksman: the Father of Antibiotics. *J Biol Chem* [Internet]. 2004 Nov 26;279(48):e7–e7. Available from: <http://www.jbc.org/content/279/48/e7.short>
 22. Schatz A, Bugle E, Waksman SA. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria. *Proc Soc Exp Biol Med* [Internet]. 1944 Jan 1;55(1):66–9. Available from: <https://journals.sagepub.com/doi/abs/10.3181/00379727-55-14461>
 23. Waksman SA, Schatz A, Reynolds DM. Production of antibiotic substances by

- actinomycetes. *Ann N Y Acad Sci.* 2010 Dec;1213(1):112–24.
24. Hutchings M, Truman A, Wilkinson B. Antibiotics: past, present and future. *Curr Opin Microbiol* [Internet]. 2019;51:72–80. Available from: <http://www.sciencedirect.com/science/article/pii/S1369527419300190>
 25. Madigan MT, Martinko JM, Stahl DA, Clark DP. Brock, *Biology of Microorganisms* [Internet]. Thirteenth. Vol. 159, Ugeskrift for Laeger. Pearson; 2013. 1041 p. Available from: <http://www.millenniumassessment.org/en/Framework.html%5Cnhttp://www.who.int/entity/globalchange/ecosystems/ecosys.pdf%5Cnhttp://www.loc.gov/catdir/toc/ecip0512/2005013229.html%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/15003161%5Cnhttp://cid.oxfordjournals.org>
 26. Hoskisson PA. *Streptomyces in Nature and Medicine: the Antibiotic Makers.* Vol. 3, Microbe Magazine. Oxford: Oxford University Press; 2008. 151–151 p.
 27. Waksman SA. Streptomycin: Background, isolation, properties, and utilization. *Science* (80-). 1953 Sep;118(3062):259–66.
 28. Kinashi H, Shimaji-Murayama M. Physical characterization of SCP1, a giant linear plasmid from *Streptomyces coelicolor*. *J Bacteriol* [Internet]. 1991 Feb;173(4):1523–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/1847368>
 29. Lin Y -S, M-Kieser H, Hopwood DA, Chen CW. The chromosomal DNA of *Streptomyces lividans* 66 is linear. Vol. 14, *Molecular Microbiology*. England; 1994. p. 1103–1103.
 30. Flärdh K, Buttner MJ. *Streptomyces* morphogenetics: Dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol.* 2009 Jan;7(1):36–49.
 31. Janssen PH. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* [Internet]. 2006 Mar;72(3):1719–28. Available from: <https://pubmed.ncbi.nlm.nih.gov/16517615>
 32. van der Meij A, Worsley SF, Hutchings MI, van Wezel GP. Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol Rev.* 2017 May;41(3):392–416.
 33. Loria R, Kers J, Joshi M. Evolution of plant pathogenicity in *Streptomyces*. *Annu Rev Phytopathol* [Internet]. 2006 Aug 8;44(1):469–87. Available from: <https://doi.org/10.1146/annurev.phyto.44.032905.091147>
 34. Bystrykh L V., Fernández-Moreno MA, Herrema JK, Malpartida F, Hopwood DA, Dijkhuizen L. Production of actinorhodin-related “blue pigments” by *Streptomyces coelicolor* A3(2). *J Bacteriol.* 1996 Apr;178(8):2238–44.
 35. GLAUERT AM, HOPWOOD DA. The fine structure of *Streptomyces coelicolor*. I. The cytoplasmic membrane system. *J Biophys Biochem Cytol* [Internet]. 1960 Jun;7(3):479–88. Available from: <https://pubmed.ncbi.nlm.nih.gov/13828388>

36. von Flaig W, Beutelspacher H, Küster E, Segler-Holzweissig G. Beiträge zur Physiologie und Morphologie der Streptomyceten. *Plant Soil* [Internet]. 1952;4(2):118–27. Available from: <https://doi.org/10.1007/BF01373641>
37. Ettlinger L, Corbaz R, Hütter R. Zur Systematik der Actinomyceten - 4. Eine Arteinteilung der Gattung *Streptomyces* Waksman et Henrici. *Arch Mikrobiol* [Internet]. 1958;31(2):326–58. Available from: <https://doi.org/10.1007/BF00446537>
38. TRESNER HD, DAVIES MC, BACKUS EJ. Electron microscopy of *Streptomyces* spore morphology and its role in species differentiation. *J Bacteriol* [Internet]. 1961 Jan;81(1):70–80. Available from: <https://pubmed.ncbi.nlm.nih.gov/13778010>
39. Kelemen GH, Brian P, Flärdh K, Chamberlin L, Chater KF, Buttner MJ. Developmental regulation of transcription of *whiE*, a locus specifying the polyketide spore pigment in *Streptomyces coelicolor* A3(2). *J Bacteriol* [Internet]. 1998 May 1;180(9):2515–21. Available from: <http://jb.asm.org/content/180/9/2515.abstract>
40. Bush MJ, Tschowri N, Schlimpert S, Flärdh K, Buttner MJ. C-di-GMP signalling and the regulation of developmental transitions in streptomycetes. *Nat Rev Microbiol* [Internet]. 2015;13(12):749–60. Available from: <https://doi.org/10.1038/nrmicro3546>
41. Gallagher KA, Schumacher MA, Bush MJ, Bibb MJ, Chandra G, Holmes NA, et al. c-di-GMP Arms an Anti- σ to Control Progression of Multicellular Differentiation in *Streptomyces*. *Mol Cell* [Internet]. 2020;77(3):586-599.e6. Available from: <http://www.sciencedirect.com/science/article/pii/S1097276519308354>
42. Jones SE, Ho L, Rees CA, Hill JE, Nodwell JR, Elliot MA. *Streptomyces* exploration is triggered by fungal interactions and volatile signals. *Elife* [Internet]. 2017 Jan 3;6:e21738. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/28044982>
43. Jones SE, Elliot MA. *Streptomyces* Exploration: Competition, Volatile Communication and New Bacterial Behaviours. *Trends Microbiol* [Internet]. 2017;25(7):522–31. Available from: <http://www.sciencedirect.com/science/article/pii/S0966842X17300239>
44. Ehrlich J, Bartz QR, Smith RM, Joslyn DA, Burkholder PR. Chloromycetin, a new antibiotic from a soil actinomycete. *Science* (80-). 1985 Oct;106(2757):417.
45. Duggar BM. Aureomycin: a Product of the Continuing Search for New Antibiotics. *Ann N Y Acad Sci*. 1948 Nov;51(2):177–81.
46. Debono M, Barnhart M, Carrell CB, Hoffmann JA, Occolowitz JL, Abbott BJ, et

- al. A21978C, a complex of new acidic peptide antibiotics: Isolation, chemistry, and mass spectral structure elucidation. *J Antibiot (Tokyo)*. 1987 Jun;40(6):761–77.
47. Tally FP, DeBruin MF. Development of daptomycin for Gram-positive infections. *J Antimicrob Chemother* [Internet]. 2000 Oct 1;46(4):523–6. Available from: <https://doi.org/10.1093/jac/46.4.523>
48. C. Campbell W. History of Avermectin and Ivermectin, with Notes on the History of Other Macrocyclic Lactone Antiparasitic Agents. *Curr Pharm Biotechnol*. 2012 May;13(6):853–65.
49. de Lima Procópio RE, da Silva IR, Martins MK, de Azevedo JL, de Araújo JM. Antibiotics produced by *Streptomyces*. *Brazilian J Infect Dis* [Internet]. 2012;16(5):466–71. Available from: <http://www.sciencedirect.com/science/article/pii/S1413867012001341>
50. Aparicio JF, Barreales EG, Payero TD, Vicente CM, de Pedro A, Santos-Aberturas J. Biotechnological production and application of the antibiotic pimarin: biosynthesis and its regulation. *Appl Microbiol Biotechnol* [Internet]. 2015/10/29. 2016 Jan;100(1):61–78. Available from: <https://pubmed.ncbi.nlm.nih.gov/26512010>
51. Dutcher JD. The discovery and development of amphotericin B. *Dis Chest*. 1968 Oct;54:Suppl 1:296-8.
52. Lyu A, Liu H, Che H, Yang L, Zhang J, Wu M, et al. Reveromycins A and B from *Streptomyces* sp. 3-10: Antifungal activity against plant pathogenic fungi in vitro and in a strawberry food model system. *Front Microbiol* [Internet]. 2017 Apr 3;8(APR):550. Available from: <https://pubmed.ncbi.nlm.nih.gov/28421050>
53. Jacob C, Weissman KJ. Unpackaging the Roles of *Streptomyces* Natural Products. *Cell Chem Biol* [Internet]. 2017;24(10):1194–5. Available from: <http://www.sciencedirect.com/science/article/pii/S2451945617303574>
54. Hollstein U. Actinomycin. Chemistry and mechanism of action. *Chem Rev* [Internet]. 1974 Dec 1;74(6):625–52. Available from: <https://pubs.acs.org/doi/abs/10.1021/cr60292a002>
55. Barona-Gómez F, Wong U, Giannakopoulos AE, Derrick PJ, Challis GL. Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. *J Am Chem Soc*. 2004 Dec;126(50):16282–3.
56. Williams K, Szwalbe AJ, Mulholland NP, Vincent JL, Bailey AM, Willis CL, et al. Heterologous Production of Fungal Maleidrides Reveals the Cryptic Cyclization Involved in their Biosynthesis. *Angew Chemie - Int Ed* [Internet]. 2016 Jun 1;55(23):6784–8. Available from: <https://doi.org/10.1002/anie.201511882>
57. Davison J, Al Fahad A, Cai M, Song Z, Yehia SY, Lazarus CM, et al. Genetic,

- molecular, and biochemical basis of fungal tropolone biosynthesis. *Proc Natl Acad Sci U S A* [Internet]. 2012 May 15;109(20):7642–7. Available from: <http://www.pnas.org/content/109/20/7642.abstract>
58. Bushley KE, Turgeon BG. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evol Biol* [Internet]. 2010;10(1):26. Available from: <https://doi.org/10.1186/1471-2148-10-26>
 59. Schor R, Cox R. Classic fungal natural products in the genomic age: The molecular legacy of Harold Raistrick. *Nat Prod Rep* [Internet]. 2018;35(3):230–56. Available from: <http://dx.doi.org/10.1039/C8NP00021B>
 60. Dreyfuss M, Härri E, Hofmann H, Kobel H, Pache W, Tscherter H. Cyclosporin A and C - New metabolites from *Trichoderma polysporum* (Link ex Pers.) Rifai. *Eur J Appl Microbiol* [Internet]. 1976;3(2):125–33. Available from: <https://doi.org/10.1007/BF00928431>
 61. Hawksworth DL, Lücking R. Fungal Diversity Revisited: 2.2 to 3.8 Million Species. In: *The Fungal Kingdom* [Internet]. American Society of Microbiology; 2017. p. 79–95. Available from: <https://www.asmscience.org/content/book/10.1128/9781555819583.chap4>
 62. Jiang ZD, An Z. Bioactive fungal natural products through classic and biocombinatorial approaches. In: Atta-ur-Rahman BT-S in NPC, editor. *Studies in Natural Products Chemistry* [Internet]. Elsevier; 2000. p. 245–72. Available from: <http://www.sciencedirect.com/science/article/pii/S1572599500800277>
 63. Charlesworth JC, Burns BP. Untapped Resources: Biotechnological Potential of Peptides and Secondary Metabolites in Archaea. *Archaea*. 2015;2015:282035.
 64. Ganz T. Defensins: Antimicrobial peptides of innate immunity. *Nat Rev Immunol* [Internet]. 2003;3(9):710–20. Available from: <https://doi.org/10.1038/nri1180>
 65. Butcher RA, Feng L, Shou Q. Discovery and biosynthesis of hybrid polyketide-nonribosomal peptides in nematodes. *FASEB J* [Internet]. 2017 Apr 1;31(S1):121.1-121.1. Available from: https://doi.org/10.1096/fasebj.31.1_supplement.121.1
 66. Fan Q, Ma J, Xu Q, Zhang J, Simion D, Carmen G, et al. Animal-derived natural products review: Focus on novel modifications and applications. *Colloids Surfaces B Biointerfaces* [Internet]. 2015;128:181–90. Available from: <http://www.sciencedirect.com/science/article/pii/S0927776515001095>
 67. Macquarrie DJ, Hardy JJE. Applications of functionalized chitosan in catalysis. *Ind Eng Chem Res* [Internet]. 2005 Nov 1;44(23):8499–520. Available from: <https://doi.org/10.1021/ie050007v>
 68. Kumari A, Yadav SK, Yadav SC. Biodegradable polymeric nanoparticles based

- drug delivery systems. *Colloids Surfaces B Biointerfaces*. 2010 Jan;75(1):1–18.
69. Liu Y, Ren L, Long K, Wang L, Wang Y. Preparation and characterization of a novel tobramycin-containing antibacterial collagen film for corneal tissue engineering. *Acta Biomater* [Internet]. 2014;10(1):289–99. Available from: <http://www.sciencedirect.com/science/article/pii/S174270611300425X>
 70. Veeresham C. Natural products derived from plants as a source of drugs. *J Adv Pharm Technol Res* [Internet]. 2012 Oct;3(4):200–1. Available from: <https://pubmed.ncbi.nlm.nih.gov/23378939>
 71. Tin-Wa M, Crane FA, Baines R, Farnsworth NR. Germination and morphine content of papaver somniferum plants produced from commercially available poppy seed. *J Pharm Sci*. 1975 Dec;64(12):2024–5.
 72. Desborough MJR, Keeling DM. The aspirin story – from willow to wonder drug. *Br J Haematol* [Internet]. 2017 Jun 1;177(5):674–83. Available from: <https://doi.org/10.1111/bjh.14520>
 73. Singh B, Sharma RA. Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. *3 Biotech* [Internet]. 2015;5(2):129–51. Available from: <https://doi.org/10.1007/s13205-014-0220-2>
 74. Kumari S, Pundhir S, Priya P, Jeena G, Punetha A, Chawla K, et al. EssOilDB: A database of essential oils reflecting terpene composition and variability in the plant kingdom. *Database* [Internet]. 2014 Dec 20;2014. Available from: <https://doi.org/10.1093/database/bau120>
 75. Schwab W, Davidovich-Rikanati R, Lewinsohn E. Biosynthesis of plant-derived flavor compounds. *Plant J* [Internet]. 2008 May 1;54(4):712–32. Available from: <https://doi.org/10.1111/j.1365-313X.2008.03446.x>
 76. C Reygaert W. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol* [Internet]. 2018 Jun 26;4(3):482–501. Available from: <https://pubmed.ncbi.nlm.nih.gov/31294229>
 77. Cox G, Wright GD. Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions. *Int J Med Microbiol*. 2013 Aug;303(6–7):287–92.
 78. Arzanlou M, Chai WC, Venter H. Intrinsic, adaptive and acquired antimicrobial resistance in Gram-negative bacteria. *Essays Biochem* [Internet]. 2017 Mar 3;61(1):49–59. Available from: <http://essays.biochemistry.org/content/61/1/49.abstract>
 79. Shallcross LJ. Editorials: Antibiotic overuse: A key driver of antimicrobial resistance. *Br J Gen Pract* [Internet]. 2014 Dec 1;64(629):604–5. Available from: <http://bjgp.org/content/64/629/604.abstract>
 80. Dcosta VM, King CE, Kalan L, Morar M, Sung WWL, Schwarz C, et al. Antibiotic resistance is ancient. *Nature*. 2011 Aug;477(7365):457–61.

81. Toth M, Smith C, Frase H, Mobashery S, Vakulenko S. An antibiotic-resistance enzyme from a deep-sea bacterium. *J Am Chem Soc.* 2010 Jan;132(2):816–23.
82. Bhullar K, Waglechner N, Pawlowski A, Koteva K, Banks ED, Johnston MD, et al. Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS One* [Internet]. 2012/04/11. 2012;7(4):e34953–e34953. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/22509370>
83. Balouiri M, Sadiki M, Ibensouda SK. Methods for in vitro evaluating antimicrobial activity: A review [Internet]. Vol. 6, *Journal of Pharmaceutical Analysis*. 2016. Available from: https://amr-review.org/sites/default/files/160525_Final paper_with cover.pdf
84. Livermore DM. The 2018 Garrod lecture: Preparing for the Black Swans of resistance. *J Antimicrob Chemother* [Internet]. 2018 Jul 25;73(11):2907–15. Available from: <https://doi.org/10.1093/jac/dky265>
85. Patel R, Piper K, Cockerill FR, Steckelberg JM, Yousten AA. The biopesticide *Paenibacillus popilliae* has a vancomycin resistance gene cluster homologous to the enterococcal VanA vancomycin resistance gene cluster. *Antimicrob Agents Chemother.* 2000 Mar;44(3):705–9.
86. Årdal C, Balasegaram M, Laxminarayan R, McAdams D, Outtersen K, Rex JH, et al. Antibiotic development — economic, regulatory and societal challenges. *Nat Rev Microbiol* [Internet]. 2020;18(5):267–74. Available from: <https://doi.org/10.1038/s41579-019-0293-3>
87. Baltz RH. Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. *J Ind Microbiol Biotechnol.* 2019 Mar;46(3–4):281–99.
88. Bode HB, Müller R. The impact of bacterial genomics on natural product research. *Angew Chemie - Int Ed* [Internet]. 2005;44(42):6828–46. Available from: http://onlinelibrary.wiley.com/store/10.1002/anie.200501080/asset/6828_ftp.pdf?v=1&t=j02ddmqf&s=fb001bd4a09994f4877260bc2d057200b3e1d12a
89. Waksman SA. Antibiotics: The duplication problem [5]. *Science* (80-) [Internet]. 1965 Mar 19;147(3664):1396–7. Available from: <http://science.sciencemag.org/content/147/3664/1396.3.abstract>
90. Rateb ME, Ebel R, Jaspars M. Natural product diversity of actinobacteria in the Atacama Desert. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol.* 2018 Aug;111(8):1467–77.
91. Millán-Aguíñaga N, Soldatou S, Brozio S, Munnoch JT, Howe J, Hoskisson PA, et al. Awakening ancient polar actinobacteria: Diversity, evolution and specialized metabolite potential. *Microbiol (United Kingdom)* [Internet].

2019;165(11):1169–80. Available from:

<https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.000845>

92. Tuttle RN, Demko AM, Patin N V., Kaponi CA, Donia MS, Dorrestein P, et al. Detection of natural products and their producers in ocean sediments. Müller V, editor. *Appl Environ Microbiol* [Internet]. 2019 Apr 15;85(8):e02830-18. Available from: <http://aem.asm.org/content/85/8/e02830-18.abstract>
93. Qin Z, Munnoch JT, Devine R, Holmes NA, Seipke RF, Wilkinson KA, et al. Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African *Tetraponera* plant-ants. *Chem Sci*. 2017 Apr;8(4):3218–27.
94. Holmes NA, Devine R, Qin Z, Seipke RF, Wilkinson B, Hutchings MI. Complete genome sequence of *Streptomyces formicae* KY5, the formicamycin producer. *J Biotechnol*. 2018 Jan;265:116–8.
95. Imai Y, Meyer KJ, Iinishi A, Favre-Godal Q, Green R, Manuse S, et al. A new antibiotic selectively kills Gram-negative pathogens. *Nature* [Internet]. 2019;576(7787):459–64. Available from: <https://doi.org/10.1038/s41586-019-1791-1>
96. Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol*. 1985;39:321–46.
97. Berdy B, Spoering AL, Ling LL, Epstein SS. In situ cultivation of previously uncultivable microorganisms using the ichip. *Nat Protoc*. 2017 Oct;12(10):2232–42.
98. Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, et al. Erratum: A new antibiotic kills pathogens without detectable resistance. *Nature* [Internet]. 2015 Jan 7;520(7547):388. Available from: <https://doi.org/10.1038/nature14098>
99. Feng Z, Chakraborty D, Dewell SB, Reddy BVB, Brady SF. Environmental DNA-encoded antibiotics fasamycins A and B inhibit FabF in type II fatty acid biosynthesis. *J Am Chem Soc*. 2012 Feb;134(6):2981–7.
100. Wang M, Carver JJ, Phelan V V., Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* [Internet]. 2016;34(8):828–37. Available from: <http://www.nature.com/nbt/journal/v34/n8/pdf/nbt.3597.pdf>
101. Risdian C, Mozef T, Wink J. Biosynthesis of polyketides in *Streptomyces*. *Microorganisms* [Internet]. 2019 May 6;7(5):124. Available from: <https://pubmed.ncbi.nlm.nih.gov/31064143>
102. Hertweck C. The biosynthetic logic of polyketide diversity. *Angew Chemie - Int*

- Ed [Internet]. 2009 Jun 15;48(26):4688–716. Available from:
<https://doi.org/10.1002/anie.200806121>
103. Staunton J, Weissman KJ. Polyketide biosynthesis: A millennium review. *Nat Prod Rep* [Internet]. 2001;18(4):380–416. Available from:
<http://pubs.rsc.org/en/Content/ArticleLanding/2001/NP/a909079g>
 104. Staunton J, Wilkinson B. Biosynthesis of Erythromycin and Rapamycin. *Chem Rev* [Internet]. 1997 Nov 1;97(7):2611–30. Available from:
<https://doi.org/10.1021/cr9600316>
 105. Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* [Internet]. 1990;348(6297):176–8. Available from: <https://doi.org/10.1038/348176a0>
 106. Ray L, Moore BS. Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Nat Prod Rep*. 2016 Feb;33(2):150–61.
 107. Das A, Khosla C. Biosynthesis of Aromatic Polyketides in Bacteria. *Acc Chem Res* [Internet]. 2009 May 19;42(5):631–9. Available from:
<https://doi.org/10.1021/ar8002249>
 108. Helfrich EJN, Lin GM, Voigt CA, Clardy J. Bacterial terpene biosynthesis: Challenges and opportunities for pathway engineering. *Beilstein J Org Chem* [Internet]. 2019 Nov 29;15:2889–906. Available from:
<https://pubmed.ncbi.nlm.nih.gov/31839835>
 109. Oldfield E, Lin FY. Terpene biosynthesis: Modularity rules. *Angew Chemie - Int Ed* [Internet]. 2011/11/21. 2012 Jan 27;51(5):1124–37. Available from:
<https://pubmed.ncbi.nlm.nih.gov/22105807>
 110. Karunanithi PS, Zerbe P. Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity [Internet]. Vol. 10, *Frontiers in Plant Science*. 2019. p. 1166. Available from:
<https://www.frontiersin.org/article/10.3389/fpls.2019.01166>
 111. Zhou F, Pichersky E. More is better: the diversity of terpene metabolism in plants. *Curr Opin Plant Biol* [Internet]. 2020;55:1–10. Available from:
<http://www.sciencedirect.com/science/article/pii/S1369526620300066>
 112. Jiang J, He X, Cane DE. Geosmin biosynthesis. *Streptomyces coelicolor* germacradienol/germacrene D synthase converts farnesyl diphosphate to geosmin. *J Am Chem Soc* [Internet]. 2006 Jun 1;128(25):8128–9. Available from: <https://doi.org/10.1021/ja062669x>
 113. Yim G, Thaker MN, Koteva K, Wright G. Glycopeptide antibiotic biosynthesis. *J Antibiot (Tokyo)* [Internet]. 2014;67(1):31–41. Available from:
<https://doi.org/10.1038/ja.2013.117>

114. Süssmuth RD, Mainz A. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angew Chemie - Int Ed* [Internet]. 2017 Mar 27;56(14):3770–821. Available from: <https://doi.org/10.1002/anie.201609079>
115. Miller BR, Gulick AM. Structural biology of nonribosomal peptide synthetases. *Methods Mol Biol* [Internet]. 2016;1401:3–29. Available from: <https://pubmed.ncbi.nlm.nih.gov/26831698>
116. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat Prod Rep* [Internet]. 2013;30(1):108–60. Available from: <http://pubs.rsc.org/en/content/articlepdf/2013/np/c2np20085f>
117. Ortega MA, Van Der Donk WA. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem Biol* [Internet]. 2016;23(1):31–44. Available from: http://ac.els-cdn.com/S2451945615004730/1-s2.0-S2451945615004730-main.pdf?_tid=ffa6da32-04c3-11e7-a7a2-00000aacb35f&acdnat=1489062731_9190eee52660a63ffb2ec2fd918e0708
118. Burkhardt BJ, Hudson GA, Dunbar KL, Mitchell DA. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat Chem Biol* [Internet]. 2015;11(8):564–70. Available from: <http://www.nature.com/nchembio/journal/v11/n8/pdf/nchembio.1856.pdf>
119. Evans RL, Latham JA, Xia Y, Klinman JP, Wilmot CM. Nuclear Magnetic Resonance Structure and Binding Studies of PqqD, a Chaperone Required in the Biosynthesis of the Bacterial Dehydrogenase Cofactor Pyrroloquinoline Quinone. *Biochemistry*. 2017;56(21):2735–46.
120. Chekan JR, Ongpipattanakul C, Nair SK. Steric complementarity directs sequence promiscuous leader binding in RiPP biosynthesis. *Proc Natl Acad Sci U S A* [Internet]. 2019 Nov 26;116(48):24049–55. Available from: <http://www.pnas.org/content/116/48/24049.abstract>
121. Crone WJK, Vior NM, Santos-Aberturas J, Schmitz LG, Leeper FJ, Truman AW. Dissecting Botromycin Biosynthesis Using Comparative Untargeted Metabolomics. *Angew Chemie - Int Ed* [Internet]. 2016;55(33):9639–43. Available from: <http://onlinelibrary.wiley.com/store/10.1002/anie.201604304/asset/anie201604304.pdf?v=1&t=iudvbjpu&s=701514ffacd5dcb68e9c7c08b152e464a6d8699f>
122. Ghodge S V., Biernat KA, Bassett SJ, Redinbo MR, Bowers AA. Post-translational claisen condensation and decarboxylation en route to the bicyclic core of pantocin A. *J Am Chem Soc* [Internet]. 2016 May 4;138(17):5487–90.

Available from: <https://doi.org/10.1021/jacs.5b13529>

123. Walton JD, Hallen-Adams HE, Luo H. Ribosomal biosynthesis of the cyclic peptide toxins of *Amanita* mushrooms. *Biopolymers* [Internet]. 2010 Jan 1;94(5):659–64. Available from: <https://doi.org/10.1002/bip.21416>
124. Santos-Aberturas J, Chandra G, Frattaruolo L, Lacret R, Pham TH, Vior NM, et al. Uncovering the unexplored diversity of thioamidated ribosomal peptides in *Actinobacteria* using the RiPPER genome mining tool. *Nucleic Acids Res* [Internet]. 2019 Mar 27;47(9):4624–37. Available from: <https://doi.org/10.1093/nar/gkz192>
125. Sivonen K, Leikoski N, Fewer DP, Jokela J. Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Appl Microbiol Biotechnol*. 2010 May;86(5):1213–25.
126. Nizet V, Beall B, Bast DJ, Datta V, Kilburn L, Low DE, et al. Genetic locus for streptolysin S production by group A streptococcus. *Infect Immun*. 2000 Jul;68(7):4245–54.
127. Asensio C, Pérez-Díaz JC, Martínez MC, Baquero F. A new family of low molecular weight antibiotics from enterobacteria. *Biochem Biophys Res Commun* [Internet]. 1976;69(1):7–14. Available from: <http://www.sciencedirect.com/science/article/pii/S0006291X76802641>
128. Heddle JG, Blance SJ, Zamble DB, Hollfelder F, Miller DA, Wentzell LM, et al. The antibiotic microcin B17 is a DNA gyrase poison: Characterisation of the mode of inhibition. *J Mol Biol*. 2001 Apr;307(5):1223–34.
129. Meteleev M, Osterman IA, Ghilarov D, Khabibullina NF, Yakimov A, Shabalin K, et al. Klebsazolicin inhibits 70S ribosome by obstructing the peptide exit tunnel. *Nat Chem Biol*. 2017;13(10):1129–36.
130. Nicolaou KC, Zak M, Rahimipour S, Estrada AA, Lee SH, O'Brate A, et al. Discovery of a biologically active thiostrepton fragment. *J Am Chem Soc*. 2005 Nov;127(43):15042–4.
131. Bhat UG, Halasi M, Gartel AL. Thiazole antibiotics target FoxM1 and induce apoptosis in human cancer cells. *PLoS One* [Internet]. 2009/05/18. 2009;4(5):e5592–e5592. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/19440351>
132. Schoof S, Pradel G, Aminake MN, Ellinger B, Baumann S, Potowski M, et al. Antiplasmodial thiostrepton derivatives: Proteasome inhibitors with a dual mode of action. *Angew Chemie - Int Ed*. 2010 Apr;49(19):3317–21.
133. Ueno M, Furukawa S, Abe F, Ushioda M, Fujine K, Johki S, et al. Suppressive effect of antibiotic siomycin on antibody production. *J Antibiot (Tokyo)*. 2004 Sep;57(9):590–6.

134. SU TL. Micrococcin, an antibacterial substance formed by a strain of. *Br J Exp Pathol* [Internet]. 1948 Oct;29(5):473–81. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/18123292>
135. Just-Baringo X, Albericio F, Álvarez M. Thiopeptide antibiotics: Retrospective and recent advances. *Mar Drugs* [Internet]. 2014 Jan 17;12(1):317–51. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/24445304>
136. CUNDLIFFE E, THOMPSON J. Concerning the Mode of Action of Micrococcin upon Bacterial Protein Synthesis. *Eur J Biochem*. 1981 Aug;118(1):47–52.
137. Rogers MJ, Cundliffe E, McCutchan TF. The antibiotic micrococcin is a potent inhibitor of growth and protein synthesis in the malaria parasite. *Antimicrob Agents Chemother* [Internet]. 1998 Mar 1;42(3):715–6. Available from: <http://aac.asm.org/content/42/3/715.abstract>
138. DUTCHER JD, VANDEPUTTE J. Thiostrepton, a new antibiotic. II. Isolation and chemical characterization. *Antibiot Annu*. 1955;3:560–1.
139. DONOVICK R, PAGANO JF, STOUT HA, WEINSTEIN MJ. Thiostrepton, a new antibiotic. I. In vitro studies. *Antibiot Annu*. 1955;3:554–9.
140. Anderson B, Hodgkin DC, Viswamitra MA. The structure of thiostrepton. *Nature* [Internet]. 1970;225(5229):233–5. Available from: <https://doi.org/10.1038/225233a0>
141. Brown LCW, Acker MG, Clardy J, Walsh CT, Fischbach MA. Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. *Proc Natl Acad Sci U S A* [Internet]. 2009/02/05. 2009 Feb 24;106(8):2549–53. Available from: <https://pubmed.ncbi.nlm.nih.gov/19196969>
142. Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, et al. Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc Natl Acad Sci U S A* [Internet]. 2005 May 17;102(20):7315–20. Available from: <http://www.pnas.org/content/102/20/7315.abstract>
143. Carroll AR, Coll JC, Bourne DJ, MacLeod JK, Zabriskie TM, Ireland CM, et al. Patellins 1-6 and trunkamide A: Novel cyclic hexa-, hepta- and octa-peptides from colonial ascidians, *Lissoclinum* sp. *Aust J Chem* [Internet]. 1996;49(6):659–67. Available from: <https://doi.org/10.1071/CH9960659>
144. Crnkovic CM, Braesel J, Krunić A, Eustáquio AS, Orjala J. Scytodecamide from the Cultured *Scytonema* sp. UIC 10036 Expands the Chemical and Genetic Diversity of Cyanobactins. *ChemBioChem* [Internet]. 2020 Nov 26;21(6):845–52. Available from: <https://doi.org/10.1002/cbic.201900511>
145. Waisvisz JM, Van Der Hoeven MG, Van Peppen J, Zwennis WCM. Bottromycin. I. A New Sulfur-containing Antibiotic. *J Am Chem Soc* [Internet]. 1957 Aug

- 1;79(16):4520–1. Available from: <https://doi.org/10.1021/ja01573a072>
146. Hou Y, Tianero MDB, Kwan JC, Wyche TP, Michel CR, Ellis GA, et al. Structure and biosynthesis of the antibiotic bottromycin D. *Org Lett* [Internet]. 2012 Oct 5;14(19):5050–3. Available from: <https://doi.org/10.1021/ol3022758>
147. Shimamura H, Gouda H, Nagai K, Hirose T, Ichioka M, Furuya Y, et al. Structure determination and total synthesis of bottromycin A2: A potent antibiotic against MRSA and VRE. *Angew Chemie - Int Ed*. 2009;48(5):914–7.
148. Hayakawa Y, Sasaki K, Nagai K, Shin-Ya K, Furihata K. Structure of Thioviridamide, a novel apoptosis inducer from *Streptomyces olivoviridis*. *J Antibiot (Tokyo)*. 2006 Jan;59(1):6–10.
149. Izawa M, Kawasaki T, Hayakawa Y. Cloning and heterologous expression of the thioviridamide biosynthesis gene cluster from *streptomyces olivoviridis*. *Appl Environ Microbiol* [Internet]. 2013 Nov 15;79(22):7110–3. Available from: <http://aem.asm.org/content/79/22/7110.abstract>
150. Montalbán-López M, Scott TA, Ramesh S, Rahman IR, van Heel AJ, Viel JH, et al. New developments in RiPP discovery, enzymology and engineering. *Nat Prod Rep* [Internet]. 2020; Available from: <http://dx.doi.org/10.1039/D0NP00027B>
151. Frattaruolo L, Lacret R, Cappello AR, Truman AW. A Genomics-Based Approach Identifies a Thioviridamide-Like Compound with Selective Anticancer Activity. *ACS Chem Biol*. 2017;12(11):2815–22.
152. Kjaerulff L, Sikandar A, Zaburannyi N, Adam S, Herrmann J, Koehnke J, et al. Thioholgamides: Thioamide-Containing Cytotoxic RiPP Natural Products. *ACS Chem Biol*. 2017;12(11):2837–41.
153. Gross E, Morell JL. The Structure of Nisin. *J Am Chem Soc* [Internet]. 1971 Sep 1;93(18):4634–5. Available from: <https://doi.org/10.1021/ja00747a073>
154. Vikeli E, Widdick DA, Batey SFD, Heine D, Holmes NA, Bibb MJ, et al. In situ activation and heterologous production of a cryptic lantibiotic from an african plant ant-derived *saccharopolyspora* species. *Appl Environ Microbiol* [Internet]. 2020 Nov 15;86(3):AEM.01876-19. Available from: <http://aem.asm.org/content/early/2019/11/11/AEM.01876-19.abstract>
155. Kloosterman AM, Cimermanic P, Elsayed SS, Du C, Hadjithomas M, Donia MS, et al. Integration of machine learning and pan-genomics expands the biosynthetic landscape of RiPP natural products. *bioRxiv* [Internet]. 2020 Jan 1;2020.05.19.104752. Available from: <http://biorxiv.org/content/early/2020/05/19/2020.05.19.104752.abstract>
156. Kawulka K, Sprules T, McKay RT, Mercier P, Diaper CM, Zuber P, et al. Structure of subtilosin A, an antimicrobial peptide from *Bacillus subtilis* with

- unusual posttranslational modifications linking cysteine sulfurs to α -carbons of phenylalanine and threonine. *J Am Chem Soc.* 2003 Apr;125(16):4726–7.
157. Lee H, Churey JJ, Worobo RW. Biosynthesis and transcriptional analysis of thurincin H, a tandem repeated bacteriocin genetic locus, produced by *Bacillus thuringiensis* SF361. *FEMS Microbiol Lett.* 2009 Oct;299(2):205–13.
 158. Balty C, Guillot A, Fradale L, Brewee C, Boulay M, Kubiak X, et al. Ruminococcin C, an anti-clostridial sactipeptide produced by a prominent member of the human microbiota *Ruminococcus gnavus*. *J Biol Chem* [Internet]. 2019 Jul 23;294(40):14512–25. Available from: <http://www.jbc.org/content/early/2019/07/23/jbc.RA119.009416.abstract>
 159. Kanki K, Kazuhiko O, Toshiaki S, Kazuro S, Hong Y, Yoko T, et al. New Antibiotic, Cypemycin Taxonomy, Fermentation, Isolation and Biological Characteristics. *J Antibiot (Tokyo)*. 1993 Nov;46(11):1666–71.
 160. Claesen J, Bibb M. Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. *Proc Natl Acad Sci U S A* [Internet]. 2010 Sep 14;107(37):16297–302. Available from: <http://www.pnas.org/content/107/37/16297.abstract>
 161. Mo T, Liu WQ, Ji W, Zhao J, Chen T, Ding W, et al. Biosynthetic Insights into Linaridin Natural Products from Genome Mining and Precursor Peptide Mutagenesis. *ACS Chem Biol.* 2017 Jun;12(6):1484–8.
 162. Claesen J, Bibb MJ. Biosynthesis and regulation of grisemycin, a new member of the linaridin family of ribosomally synthesized peptides produced by *Streptomyces griseus* IFO 13350. *J Bacteriol* [Internet]. 2011 May 15;193(10):2510–6. Available from: <http://jb.asm.org/content/193/10/2510.abstract>
 163. Rateb ME, Zhai Y, Ehrner E, Rath CM, Wang X, Tabudravu J, et al. Legonaridin, a new member of linaridin RiPP from a Ghanaian *Streptomyces* isolate. *Org Biomol Chem* [Internet]. 2015;13(37):9585–92. Available from: <http://pubs.rsc.org/en/Content/ArticleLanding/2015/OB/C5OB01269D>
 164. Shang Z, Winter JM, Kauffman CA, Yang I, Fenical W. Salinipeptins: Integrated Genomic and Chemical Approaches Reveal Unusual d -Amino Acid-Containing Ribosomally Synthesized and Post-Translationally Modified Peptides (RiPPs) from a Great Salt Lake *Streptomyces* sp. *ACS Chem Biol* [Internet]. 2019 Mar 15;14(3):415–25. Available from: <https://doi.org/10.1021/acscchembio.8b01058>
 165. Ortiz-López FJ, Carretero-Molina D, Sánchez-Hidalgo M, Martín J, González I, Román-Hurtado F, et al. Cacaoidin, First Member of the New Lanthidin RiPP Family. *Angew Chemie - Int Ed* [Internet]. 2020 May 14;59(31):12654–8. Available from: <https://doi.org/10.1002/anie.202005187>

166. Wilson KA, Kalkum M, Ottesen J, Yuzenkova J, Chait BT, Landick R, et al. Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail. *J Am Chem Soc* [Internet]. 2003 Oct 1;125(41):12475–83. Available from: <https://doi.org/10.1021/ja036756q>
167. Mukhopadhyay J, Sineva E, Knight J, Levy RM, Ebright RH. Antibacterial peptide Microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. *Mol Cell* [Internet]. 2004 Jun 18;14(6):739–51. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/15200952>
168. Salomon RA, Farias RN. Microcin 25, a novel antimicrobial peptide produced by *Escherichia coli*. *J Bacteriol* [Internet]. 1992 Nov;174(22):7428–35. Available from: <https://pubmed.ncbi.nlm.nih.gov/1429464>
169. Blond A, Péduzzi J, Goulard C, Chiuchiolo MJ, Barthélémy M, Prigent Y, et al. The cyclic structure of microcin J25, a 21-residue peptide antibiotic from *Escherichia coli*. *Eur J Biochem*. 1999 Feb;259(3):747–56.
170. Rosengren KJ, Clark RJ, Daly NL, Göransson U, Jones A, Craik DJ. Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. *J Am Chem Soc*. 2003 Oct;125(41):12464–74.
171. Cheung-Lee WL, Parry ME, Zong C, Cartagena AJ, Darst SA, Connell ND, et al. Discovery of Ubonodin, an Antimicrobial Lasso Peptide Active against Members of the *Burkholderia cepacia* Complex. *ChemBioChem* [Internet]. 2020 May 4;21(9):1335–40. Available from: <https://doi.org/10.1002/cbic.201900707>
172. Cheung-Lee WL, Cao L, Link AJ. Pandonodin: A Proteobacterial Lasso Peptide with an Exceptionally Long C-Terminal Tail. *ACS Chem Biol* [Internet]. 2019 Dec 20;14(12):2783–92. Available from: <https://doi.org/10.1021/acscchembio.9b00676>
173. Ibrahim M, Guillot A, Wessner F, Algaron F, Besset C, Courtin P, et al. Control of the transcription of a short gene encoding a cyclic peptide in *Streptococcus thermophilus*: A new quorum-sensing system? *J Bacteriol* [Internet]. 2007/10/05. 2007 Dec;189(24):8845–54. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/17921293>
174. Schramma KR, Bushin LB, Seyedsayamdost MR. Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nat Chem* [Internet]. 2015;7(5):431–7. Available from: <http://www.nature.com/nchem/journal/v7/n5/pdf/nchem.2237.pdf>
175. Isley NA, Endo Y, Wu ZC, Covington BC, Bushin LB, Seyedsayamdost MR, et al. Total Synthesis and Stereochemical Assignment of Streptide. *J Am Chem Soc* [Internet]. 2019 Oct 30;141(43):17361–9. Available from: <https://doi.org/10.1021/jacs.9b09067>

176. Hamada T, Sugawara T, Matsunaga S, Fusetani N. Polytheonamides, unprecedented highly cytotoxic polypeptides, from the marine sponge theonella swinhoei. 1. Isolation and component amino acids. *Tetrahedron Lett* [Internet]. 1994;35(5):719–20. Available from: <http://www.sciencedirect.com/science/article/pii/S0040403900757996>
177. Freeman MF, Gurgui C, Helf MJ, Morinaka BI, Uria AR, Oldham NJ, et al. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science* (80-) [Internet]. 2012;338(6105):387–90. Available from: <http://science.sciencemag.org/content/sci/338/6105/387.full.pdf>
178. Fuchs SW, Lackner G, Morinaka BI, Morishita Y, Asai T, Riniker S, et al. A Lanthipeptide-like N-Terminal Leader Region Guides Peptide Epimerization by Radical SAM Epimerases: Implications for RiPP Evolution. *Angew Chemie - Int Ed* [Internet]. 2016;55(40):12330–3. Available from: <http://onlinelibrary.wiley.com/store/10.1002/anie.201602863/asset/anie201602863.pdf?v=1&t=iudvbcgv&s=efd901b883c9561352c07610dbeb3b518202e861>
179. Hamada T, Matsunaga S, Fujiwara M, Fujita K, Hirota H, Schmucki R, et al. Solution structure of polytheonamide B, a highly cytotoxic nonribosomal polypeptide from marine sponge. *J Am Chem Soc* [Internet]. 2010 Sep 22;132(37):12941–5. Available from: <https://doi.org/10.1021/ja104616z>
180. Anthony C. The pyrroloquinoline quinone (PQQ)-containing quinoprotein dehydrogenases. *Biochem Soc Trans*. 1998 Aug;26(3):413–7.
181. Hallen HE, Luo H, Scott-Craig JS, Walton JD. Gene family encoding the major toxins of lethal *Amanita* mushrooms. *Proc Natl Acad Sci U S A* [Internet]. 2007 Nov 27;104(48):19097–101. Available from: <http://www.pnas.org/content/104/48/19097.abstract>
182. Luo H, Hallen-Adams HE, Scott-Craig JS, Walton JD. Ribosomal biosynthesis of α -amanitin in *Galerina marginata*. *Fungal Genet Biol* [Internet]. 2012;49(2):123–9. Available from: <http://www.sciencedirect.com/science/article/pii/S1087184511002040>
183. Vogt E, Künzler M. Discovery of novel fungal RiPP biosynthetic pathways and their application for the development of peptide therapeutics. *Appl Microbiol Biotechnol* [Internet]. 2019;103(14):5567–81. Available from: <https://doi.org/10.1007/s00253-019-09893-x>
184. May JP, Perrin DM. Tryptathionine bridges in peptide synthesis. *Biopolym - Pept Sci Sect* [Internet]. 2007;88(5):714–24. Available from: <https://doi.org/10.1002/bip.20807>
185. Ding W, Liu WQ, Jia Y, Li Y, Van Der Donk WA, Zhang Q. Biosynthetic investigation of phomopsins reveals a widespread pathway for ribosomal natural

- products in Ascomycetes. *Proc Natl Acad Sci U S A* [Internet]. 2016/03/15. 2016 Mar 29;113(13):3521–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/26979951>
186. Van Der Velden NS, Kälin N, Helf MJ, Piel J, Freeman MF, Künzler M. Autocatalytic backbone N-methylation in a family of ribosomal peptide natural products. *Nat Chem Biol* [Internet]. 2017;13(8):833–5. Available from: <https://doi.org/10.1038/nchembio.2393>
 187. Ramm S, Krawczyk B, Mühlenweg A, Poch A, Mösker E, Süssmuth RD. A Self-Sacrificing N-Methyltransferase Is the Precursor of the Fungal Natural Product Omphalotin. *Angew Chemie - Int Ed*. 2017;56(33):9994–7.
 188. Aldemir H, Gulder TAM. Expanding the Structural Space of Ribosomal Peptides: Autocatalytic N-Methylation in Omphalotin Biosynthesis. *Angew Chemie - Int Ed* [Internet]. 2017 Oct 23;56(44):13570–2. Available from: <https://doi.org/10.1002/anie.201708456>
 189. Johnson RD, Lane GA, Koulman A, Cao M, Fraser K, Fleetwood DJ, et al. A novel family of cyclic oligopeptides derived from ribosomal peptide synthesis of an in planta-induced gene, *gigA*, in *Epichloë* endophytes of grasses. *Fungal Genet Biol*. 2015 Dec;85:14–24.
 190. Craik DJ, Daly NL, Bond T, Waine C. Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J Mol Biol* [Internet]. 1999;294(5):1327–36. Available from: <http://www.sciencedirect.com/science/article/pii/S0022283699933831>
 191. Franke B, Jayasena AS, Fisher MF, Swedberg JE, Taylor NL, Mylne JS, et al. Diverse cyclic seed peptides in the Mexican zinnia (*Zinnia haageana*). *Biopolymers*. 2016 Nov;106(6):806–17.
 192. Shim YY, Song Z, Jadhav PD, Reaney MJT. Orbitides from flaxseed (*Linum usitatissimum* L.): A comprehensive review. *Trends Food Sci Technol* [Internet]. 2019;93:197–211. Available from: <http://www.sciencedirect.com/science/article/pii/S0924224419303115>
 193. Kersten RD, Weng JK. Gene-guided discovery and engineering of branched cyclic peptides in plants. *Proc Natl Acad Sci U S A* [Internet]. 2018 Nov 13;115(46):E10961–9. Available from: <http://www.pnas.org/content/115/46/E10961.abstract>
 194. Yahara S, Shigeyama C, Nohara T, Okuda H, Wakamatsu K, Yasuhara T. Structures of anti-ace and -renin peptides from *lycii radices* cortex. *Tetrahedron Lett* [Internet]. 1989;30(44):6041–2. Available from: <http://www.sciencedirect.com/science/article/pii/S0040403901938493>
 195. Dunbar KL, Tietz JI, Cox CL, Burkhart BJ, Mitchell DA. Identification of an

- Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products. *J Am Chem Soc.* 2015;137(24):7672–7.
196. Dunbar KL, Chekan JR, Cox CL, Burkhart BJ, Nair SK, Mitchell DA. Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nat Chem Biol.* 2014;10(10):823–9.
 197. Burkhart BJ, Schwalen CJ, Mann G, Naismith JH, Mitchell DA. YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. *Chem Rev.* 2017;117(8):5389–456.
 198. Dunbar KL, Melby JO, Mitchell DA. YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat Chem Biol.* 2012;8(6):569–75.
 199. Koehnke J, Bent AF, Zollman D, Smith K, Houssen WE, Zhu X, et al. The cyanobactin heterocyclase enzyme: A processive adenylase that operates with a defined order of reaction. *Angew Chemie - Int Ed [Internet].* 2013/11/08. 2013 Dec 23;52(52):13991–6. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/24214017>
 200. Dunbar KL, Mitchell DA. Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives. *J Am Chem Soc.* 2013 Jun;135(23):8692–701.
 201. Melby JO, Nard NJ, Mitchell DA. Thiazole/oxazole-modified microcins: Complex natural products from ribosomal templates. *Curr Opin Chem Biol [Internet].* 2011;15(3):369–78. Available from: http://ac.els-cdn.com/S1367593111000366/1-s2.0-S1367593111000366-main.pdf?_tid=1e985292-9488-11e6-af08-00000aab0f27&acdnat=1476722483_8fb3e05ca097f7a35350e927fea68609
 202. Li YM, Milne JC, Madison LL, Kolter R, Walsh CT. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: Microcin B17 synthase. *Science (80-).* 1996 Nov;274(5290):1188–93.
 203. Mahanta N, Szantai-Kis DM, Petersson EJ, Mitchell DA. Biosynthesis and Chemical Applications of Thioamides. *ACS Chem Biol [Internet].* 2019 Feb 15;14(2):142–63. Available from: <https://doi.org/10.1021/acschembio.8b01022>
 204. Franz L, Adam S, Santos-Aberturas J, Truman AW, Koehnke J. Macroamidine Formation in Botromycins Is Catalyzed by a Divergent YcaO Enzyme. *J Am Chem Soc.* 2017;139(50):18158–61.
 205. Travin DY, Metelev M, Serebryakova M, Komarova ES, Osterman IA, Ghilarov D, et al. Biosynthesis of Translation Inhibitor Klebsazolicin Proceeds through Heterocyclization and N-Terminal Amidine Formation Catalyzed by a Single YcaO Enzyme. *J Am Chem Soc.* 2018;140(16):5625–33.
 206. Nayak DD, Mahanta N, Mitchell DA, Metcalf WW. Post-translational

- thioamidation of methyl-coenzyme M reductase, a key enzyme in methanogenic and methanotrophic archaea. *Elife*. 2017;6:18.
207. Liu J, Lin Z, Li Y, Zheng Q, Chen D, Liu W. Insights into the thioamidation of thiopeptins to enhance the understanding of the biosynthetic logic of thioamide-containing thiopeptides. *Org Biomol Chem* [Internet]. 2019;17(15):3727–31. Available from: <http://dx.doi.org/10.1039/C9OB00402E>
 208. Tang J, Lu J, Luo Q, Wang H. Discovery and biosynthesis of thioviridamide-like compounds. *Chinese Chem Lett* [Internet]. 2018;29(7):1022–8. Available from: <http://www.sciencedirect.com/science/article/pii/S1001841718301918>
 209. Kenney GE, Dassama LMK, Pandelia ME, Gizzi AS, Martinie RJ, Gao P, et al. The biosynthesis of methanobactin. *Science* (80-) [Internet]. 2018 Mar 23;359(6382):1411–6. Available from: <http://science.sciencemag.org/content/359/6382/1411.abstract>
 210. Mahanta N, Hudson GA, Mitchell DA. Radical S-Adenosylmethionine Enzymes Involved in RiPP Biosynthesis. *Biochemistry*. 2017;56(40):5229–44.
 211. Benjdia A, Pierre S, Gherasim C, Guillot A, Carmona M, Amara P, et al. The thiostrepton A tryptophan methyltransferase TsrM catalyses a cob(II)alamin-dependent methyl transfer reaction. *Nat Commun*. 2015 Oct;6:8377.
 212. Parent A, Guillot A, Benjdia A, Chartier G, Leprince J, Berteau O. The B12-Radical SAM Enzyme PoyC Catalyzes Valine C β -Methylation during Polytheonamide Biosynthesis. *J Am Chem Soc* [Internet]. 2016 Dec 7;138(48):15515–8. Available from: <https://doi.org/10.1021/jacs.6b06697>
 213. Liao R, Duan L, Lei C, Pan H, Ding Y, Zhang Q, et al. Thiopeptide Biosynthesis Featuring Ribosomally Synthesized Precursor Peptides and Conserved Posttranslational Modifications. *Chem Biol* [Internet]. 2009 Feb 27;16(2):141–7. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2676563/>
 214. Zhang Q, Chen D, Lin J, Liao R, Tong W, Xu Z, et al. Characterization of NocL involved in thiopeptide nocathiacin I biosynthesis: A [4Fe-4S] cluster and the catalysis of a radical S-adenosylmethionine enzyme. *J Biol Chem*. 2011 Jun;286(24):21287–94.
 215. Mahanta N, Zhang Z, Hudson GA, Van Der Donk WA, Mitchell DA. Reconstitution and Substrate Specificity of the Radical S-Adenosyl-methionine Thiazole C-Methyltransferase in Thiomuracin Biosynthesis. *J Am Chem Soc*. 2017 Mar;139(12):4310–3.
 216. Benjdia A, Balty C, Berteau O. Radical SAM enzymes in the biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPPs). *Front Chem*. 2017;5(NOV).
 217. Ding W, Mo T, Mandalapu D, Zhang Q. Substrate specificity of the cypemycin

- decarboxylase CypD. *Synth Syst Biotechnol* [Internet]. 2018 Sep 15;3(3):159–62. Available from: <https://pubmed.ncbi.nlm.nih.gov/30345401>
218. Liscombe DK, Louie G V., Noel JP. Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat Prod Rep* [Internet]. 2012;29(10):1238–50. Available from: <http://dx.doi.org/10.1039/C2NP20029E>
 219. Truman AW. Cyclisation mechanisms in the biosynthesis of ribosomally synthesised and post-translationally modified peptides. *Beilstein J Org Chem* [Internet]. 2016;12:1250–68. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4979651/pdf/Beilstein_J_Org_Chem-12-1250.pdf
 220. Zhang Q, Yu Y, Vélasquez JE, Van Der Donk WA. Evolution of lanthipeptide synthetases. *Proc Natl Acad Sci U S A* [Internet]. 2012/10/15. 2012 Nov 6;109(45):18361–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/23071302>
 221. Müller WM, Schmiederer T, Ensle P, Süssmuth RD. In vitro biosynthesis of the prepeptide of type-III lantibiotic labyrinthopeptin A2 including formation of a C-C bond as a post-translational modification. *Angew Chemie - Int Ed*. 2010 Mar;49(13):2436–40.
 222. Pag U, Sahl H-G. Multiple Activities in Lantibiotics - Models for the Design of Novel Antibiotics? *Curr Pharm Des*. 2005;8(9):815–33.
 223. Kodani S, Hudson ME, Durrant MC, Buttner MJ, Nodwell JR, Willey JM. The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene ramS in *Streptomyces coelicolor*. *Proc Natl Acad Sci U S A* [Internet]. 2004 Aug 3;101(31):11448–53. Available from: <http://www.pnas.org/content/101/31/11448.abstract>
 224. Imperiali B, Fisher SL, Moats RA, Prins TJ. A Conformational Study of Peptides with the General Structure Ac-I-Xaa-Pro-d-Xaa-I-Xaa-NH₂: Spectroscopic Evidence for a Peptide with Significant β -Turn Character in Water and in Dimethyl Sulfoxide. *J Am Chem Soc* [Internet]. 1992 Apr 1;114(9):3182–8. Available from: <https://doi.org/10.1021/ja00035a002>
 225. Cotter PD, O'Connor PM, Draper LA, Lawton EM, Deegan LH, Hill C, et al. Posttranslational conversion of L-serines to D-alanines is vital for optimal production and activity of the lantibiotic lactacin 3147. *Proc Natl Acad Sci U S A* [Internet]. 2005 Dec 20;102(51):18584–9. Available from: <http://www.pnas.org/content/102/51/18584.abstract>
 226. Rink R, Arkema-Meter A, Baudoin I, Post E, Kuipers A, Nelemans SA, et al. To protect peptide pharmaceuticals against peptidases. *J Pharmacol Toxicol Methods* [Internet]. 2010;61(2):210–8. Available from: <http://www.sciencedirect.com/science/article/pii/S1056871910000262>

227. Heck SD, Faraci WS, Kelbaugh PR, Saccomano NA, Thadeio PF, Volkmann RA. Posttranslational amino acid epimerization: Enzyme-catalyzed isomerization of amino acid residues in peptide chains. *Proc Natl Acad Sci U S A* [Internet]. 1996 Apr 30;93(9):4036–9. Available from: <http://www.pnas.org/content/93/9/4036.abstract>
228. Morinaka BI, Vagstad AL, Helf MJ, Gugger M, Kegler C, Freeman MF, et al. Radical S-adenosyl methionine epimerases: Regioselective introduction of diverse D -amino acid patterns into peptide natural products. *Angew Chemie - Int Ed* [Internet]. 2014 Aug 4;53(32):8503–7. Available from: <https://doi.org/10.1002/anie.201400478>
229. Skaugen M, Nissen-Meyer J, Jung G, Stevanovic S, Sletten K, Abildgaard CIM, et al. In vivo conversion of L-serine to D-alanine in a ribosomally synthesized polypeptide. *J Biol Chem*. 1994 Nov;269(44):27183–5.
230. Repka LM, Chekan JR, Nair SK, Van Der Donk WA. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chem Rev* [Internet]. 2017 Apr 26;117(8):5457–520. Available from: <https://doi.org/10.1021/acs.chemrev.6b00591>
231. Ryan MP, Jack RW, Josten M, Sahl HG, Jung G, Ross RP, et al. Extensive post-translational modification, including serine to D-alanine conversion, in the two-component lantibiotic, lactacin 3147. *J Biol Chem*. 1999 Dec;274(53):37544–50.
232. Freeman MF, Helf MJ, Bhushan A, Morinaka BI, Piel J. Seven enzymes create extraordinary molecular complexity in an uncultivated bacterium. *Nat Chem* [Internet]. 2017;9(4):387–95. Available from: <https://doi.org/10.1038/nchem.2666>
233. Frey PA, Hegeman AD, Ruzicka FJ. The radical SAM superfamily. *Crit Rev Biochem Mol Biol*. 2008;43(1):63–88.
234. Koehnke J, Bent A, Houssen WE, Zollman D, Morawitz F, Shirran S, et al. The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nat Struct Mol Biol* [Internet]. 2012;19(8):767–72. Available from: <https://doi.org/10.1038/nsmb.2340>
235. Sikandar A, Franz L, Adam S, Santos-Aberturas J, Horbal L, Luzhetskyy A, et al. The bottromycin epimerase BotH defines a group of atypical α/β -hydrolase-fold enzymes. *Nat Chem Biol* [Internet]. 2020;16(9):1013–8. Available from: <https://doi.org/10.1038/s41589-020-0569-y>
236. Ghilarov D, Serebryakova M, Stevenson CEM, Hearnshaw SJ, Volkov DS, Maxwell A, et al. The Origins of Specificity in the Microcin-Processing Protease TldD/E. *Structure*. 2017 Oct;25(10):1549-1561.e5.
237. Huo L, Rachid S, Stadler M, Wenzel SC, Müller R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem Biol*. 2012

Oct;19(10):1278–87.

238. Hudson GA, Zhang Z, Tietz JI, Mitchell DA, Van Der Donk WA. In Vitro Biosynthesis of the Core Scaffold of the Thiopeptide Thiomuracin. *J Am Chem Soc* [Internet]. 2015 Dec 30;137(51):16012–5. Available from: <https://doi.org/10.1021/jacs.5b10194>
239. Rince A, Dufour A, Le Pogam S, Thuault D, Bourgeois CM, Le Pennec JP. Cloning, expression, and nucleotide sequence of genes involved in production of lactococcin DR, a bacteriocin from *Lactococcus lactis* subsp. *lactis*. *Appl Environ Microbiol*. 1994 May;60(5):1652–7.
240. Havarstein LS, Diep DB, Nes IF. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol Microbiol* [Internet]. 1995 Apr 1;16(2):229–40. Available from: <https://doi.org/10.1111/j.1365-2958.1995.tb02295.x>
241. Widdick DA, Dodd HM, Barraille P, White J, Stein TH, Chater KF, et al. Cloning and engineering of the cinnamycin biosynthetic gene cluster from *Streptomyces cinnamoneus cinnamoneus* DSM 40005. *Proc Natl Acad Sci U S A* [Internet]. 2003 Apr 1;100(7):4316–21. Available from: <http://www.pnas.org/content/100/7/4316.abstract>
242. Zhang Q, Doroghazi JR, Zhao X, Walker MC, van der Donk WA. Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in Actinobacteria. Elliot MA, editor. *Appl Environ Microbiol* [Internet]. 2015 Jul 1;81(13):4339–50. Available from: <http://aem.asm.org/content/81/13/4339.abstract>
243. Völler GH, Krawczyk JM, Pesic A, Krawczyk B, Nachtigall J, Süssmuth RD. Characterization of New Class III Lantibiotics-Erythraeptin, Avermipeptin and Griseopeptin from *Saccharopolyspora erythraea*, *Streptomyces avermitilis* and *Streptomyces griseus* Demonstrates Stepwise N-Terminal Leader Processing. *ChemBioChem* [Internet]. 2012 May 29;13(8):1174–83. Available from: <https://doi.org/10.1002/cbic.201200118>
244. Völler GH, Krawczyk B, Ensle P, Süssmuth RD. Involvement and unusual substrate specificity of a prolyl oligopeptidase in class III lanthipeptide maturation. *J Am Chem Soc* [Internet]. 2013 May 22;135(20):7426–9. Available from: <https://doi.org/10.1021/ja402296m>
245. Chen S, Xu B, Chen E, Wang J, Lu J, Donadio S, et al. Zn-dependent bifunctional proteases are responsible for leader peptide processing of class III lanthipeptides. *Proc Natl Acad Sci U S A* [Internet]. 2019 Feb 12;116(7):2533–8. Available from: <http://www.pnas.org/content/116/7/2533.abstract>
246. Gruber CW, Čemažar M, Clark RJ, Horibe T, Renda RF, Anderson MA, et al. A

- novel plant protein-disulfide isomerase involved in the oxidative folding of cystine knot defense proteins. *J Biol Chem*. 2007 Jul;282(28):20435–46.
247. Saska I, Gillon AD, Hatsugai N, Dietzgen RG, Hara-Nishimura I, Anderson MA, et al. An asparaginyl endopeptidase mediates in vivo protein backbone cyclization. *J Biol Chem*. 2007 Oct;282(40):29721–8.
 248. Condie JA, Nowak G, Reed DW, Balsevich JJ, Reaney MJT, Arnison PG, et al. The biosynthesis of Caryophyllaceae-like cyclic peptides in *Saponaria vaccaria* L. from DNA-encoded precursors. *Plant J*. 2011 Aug;67(4):682–90.
 249. Chekan JR, Estrada P, Covello PS, Nair SK. Characterization of the macrocyclase involved in the biosynthesis of RiPP cyclic peptides in plants. *Proc Natl Acad Sci U S A*. 2017;114(25):6551–6.
 250. Ding X, Hou X, Xie K, Xiong L. Genome-wide identification of BURP domain-containing genes in rice reveals a gene family with diverse structures and responses to abiotic stresses. *Planta*. 2009 Jun;230(1):149–63.
 251. Schwalen CJ, Hudson GA, Kille B, Mitchell DA. Bioinformatic Expansion and Discovery of Thiopeptide Antibiotics. *J Am Chem Soc* [Internet]. 2018 Aug 1;140(30):9494–501. Available from: <https://doi.org/10.1021/jacs.8b03896>
 252. Skinnider MA, Johnston CW, Edgar RE, Dejong CA, Merwin NJ, Rees PN, et al. Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *Proc Natl Acad Sci U S A*. 2016;113(42):E6343–51.
 253. Tietz JI, Schwalen CJ, Patel PS, Maxson T, Blair PM, Tai HC, et al. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat Chem Biol*. 2017;13(5):470–8.
 254. Russell AH, Truman AW. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Comput Struct Biotechnol J* [Internet]. 2020;18:1838–51. Available from: <http://www.sciencedirect.com/science/article/pii/S2001037020303214>
 255. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* (80-). 1995 Jul;269(5223):496–512.
 256. Turk DC. The pathogenicity of *Haemophilus influenzae*. *J Med Microbiol*. 1984 Aug;18(1):1–16.
 257. Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* [Internet]. 2002;417(6885):141–7. Available from: <http://www.nature.com/nature/journal/v417/n6885/pdf/417141a.pdf>
 258. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, et al. Complete genome sequence and comparative analysis of the industrial

- microorganism *Streptomyces avermitilis*. *Nat Biotechnol*. 2003 May;21(5):526–31.
259. Nett M, Ikeda H, Moore BS. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* [Internet]. 2009;26(11):1362–84. Available from:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3063060/pdf/nihms-172646.pdf>
 260. Baltz RH. Gifted microbes for genome mining and natural product discovery. *J Ind Microbiol Biotechnol*. 2017 May;44(4–5):573–88.
 261. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*. 2012 Feb;28(4):464–9.
 262. Lu G, Moriyama EN. Vector NTI, a balanced all-in-one sequence analysis suite. *Brief Bioinform* [Internet]. 2004 Dec 1;5(4):378–88. Available from:
<https://doi.org/10.1093/bib/5.4.378>
 263. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* [Internet]. 2012/04/27. 2012 Jun 15;28(12):1647–9. Available from:
<https://pubmed.ncbi.nlm.nih.gov/22543367>
 264. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792–7.
 265. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994 Nov;22(22):4673–80.
 266. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990 Oct;215(3):403–10.
 267. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* [Internet]. 2015;10(6):845–58. Available from: <https://doi.org/10.1038/nprot.2015.053>
 268. Expasy. Expasy ProtParam tool [Internet]. SIB Swiss Institution of Bioinformatics. 2019. Available from: <https://web.expasy.org/protparam/>
 269. de Jong A, van Hijum SAFT, Bijlsma JJE, Kok J, Kuipers OP. BAGEL: A web-based bacteriocin genome mining tool. *Nucleic Acids Res* [Internet]. 2006 Jul 1;34(WEB. SERV. ISS.):W273–9. Available from:
<https://www.ncbi.nlm.nih.gov/pubmed/16845009>
 270. de Jong A, van Heel AJ, Kok J, Kuipers OP. BAGEL2: Mining for bacteriocins in genomic data. *Nucleic Acids Res* [Internet]. 2010/05/12. 2010 Jul;38(SUPPL.

- 2):W647–51. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/20462861>
271. van Heel AJ, de Jong A, Montalbán-López M, Kok J, Kuipers OP. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res* [Internet]. 2013/05/15. 2013 Jul;41(Web Server issue):W448–53. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23677608>
 272. Van Heel AJ, De Jong A, Song C, Viel JH, Kok J, Kuipers OP. BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res*. 2018 Jul;46(W1):W278–81.
 273. Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, et al. AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* [Internet]. 2011/06/14. 2011 Jul;39(SUPPL. 2):W339–46. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/21672958>
 274. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, et al. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* [Internet]. 2013/06/03. 2013 Jul;41(Web Server issue):W204–12. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23737449>
 275. Weber T, Blin K, Duddela S, Krug D, Kim HU, Brucoleri R, et al. AntiSMASH 3.0-A comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* [Internet]. 2015/05/06. 2015 Jul 1;43(W1):W237–43. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/25948579>
 276. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, et al. AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* [Internet]. 2017 Jul 3;45(W1):W36–41. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/28460038>
 277. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* [Internet]. 2019 Apr 29;47(W1):W81–7. Available from: <https://doi.org/10.1093/nar/gkz310>
 278. Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH. PlantiSMASH: Automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res*. 2017 Jul;45(W1):W55–63.
 279. Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster ALH, et al. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res* [Internet]. 2015/10/05. 2015 Nov 16;43(20):9645–62. Available from:

<https://www.ncbi.nlm.nih.gov/pubmed/26442528>

280. Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA. PRISM 3: Expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res* [Internet]. 2017 Apr 28;45(W1):W49–54. Available from: <https://doi.org/10.1093/nar/gkx320>
281. Agrawal P, Khater S, Gupta M, Sain N, Mohanty D. RiPPMiner: A bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res*. 2017;45(W1):W80–8.
282. Bouslimani A, Sanchez LM, Garg N, Dorrestein PC. Mass spectrometry of natural products: Current, emerging and future technologies. *Nat Prod Rep* [Internet]. 2014;31(6):718–29. Available from: <http://dx.doi.org/10.1039/C4NP00044G>
283. Kersten RD, Yang YL, Xu Y, Cimerancic P, Nam SJ, Fenical W, et al. A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol* [Internet]. 2011;7(11):794–802. Available from: <https://doi.org/10.1038/nchembio.684>
284. Mohimani H, Kersten RD, Liu WT, Wang M, Purvine SO, Wu S, et al. Automated genome mining of ribosomal peptide natural products. *ACS Chem Biol* [Internet]. 2014;9(7):1545–51. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4215869/pdf/cb500199h.pdf>
285. Cao L, Gurevich A, Alexander KL, Naman CB, Leão T, Glukhov E, et al. MetaMiner: A Scalable Peptidogenomics Approach for Discovery of Ribosomal Peptide Natural Products with Blind Modifications from Microbial Communities. *Cell Syst* [Internet]. 2019;9(6):600-608.e4. Available from: <http://www.sciencedirect.com/science/article/pii/S2405471219303126>
286. Morton JT, Freed SD, Lee SW, Friedberg I. A large scale prediction of bacteriocin gene blocks suggests a wide functional spectrum for bacteriocins. *BMC Bioinformatics* [Internet]. 2015;16(1):381. Available from: <https://doi.org/10.1186/s12859-015-0792-9>
287. Behsaz B, Mohimani H, Gurevich A, Prjibelski A, Fisher M, Vargas F, et al. De Novo Peptide Sequencing Reveals Many Cyclopeptides in the Human Gut and Other Environments. *Cell Syst* [Internet]. 2020;10(1):99-108.e5. Available from: <http://www.sciencedirect.com/science/article/pii/S240547121930393X>
288. Mohimani H, Gurevich A, Mikheenko A, Garg N, Nothias LF, Ninomiya A, et al. Dereplication of peptidic natural products through database search of mass spectra. *Nat Chem Biol*. 2017 Jan;13(1):30–7.
289. Gurevich A, Mikheenko A, Shlemov A, Korobeynikov A, Mohimani H, Pevzner PA. Increased diversity of peptidic natural products revealed by modification-

- tolerant database search of mass spectra. *Nat Microbiol* [Internet]. 2018;3(3):319–27. Available from: <https://doi.org/10.1038/s41564-017-0094-2>
290. Merwin NJ, Mousa WK, Dejong CA, Skinnider MA, Cannon MJ, Li H, et al. DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc Natl Acad Sci U S A* [Internet]. 2020 Jan 7;117(1):371–80. Available from: <http://www.pnas.org/content/117/1/371.abstract>
 291. Hudson GA, Burkhardt BJ, DiCaprio AJ, Schwalen CJ, Kille B, Pogorelov T V., et al. Bioinformatic mapping of radical S-adenosylmethionine-dependent ribosomally synthesized and post-translationally modified peptides identifies new C α , C β , and C γ -linked thioether-containing peptides. *J Am Chem Soc*. 2019 May;141(20):8228–38.
 292. de los Santos ELC. NeuRiPP: Neural network identification of RiPP precursor peptides. *Sci Rep* [Internet]. 2019;9(1):13406. Available from: <https://doi.org/10.1038/s41598-019-49764-z>
 293. Li J, Qu X, He X, Duan L, Wu G, Bi D, et al. ThioFinder: A Web-Based Tool for the Identification of Thiopeptide Gene Clusters in DNA Sequences. *PLoS One* [Internet]. 2012/09/24. 2012;7(9):e45878–e45878. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23029291>
 294. Kloosterman A, Shelton K, van Wezel G, Medema M, Mitchell D. RRE-Finder: A Genome-Mining Tool for Class-Independent RiPP Discovery. *mSystems*. 2020 Sep 1;2020.03.14.992123.
 295. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* [Internet]. 2005 Jul 1;33(SUPPL. 2):W244–8. Available from: <https://pubmed.ncbi.nlm.nih.gov/15980461>
 296. Iftime D, Jasyk M, Kulik A, Imhoff JF, Stegmann E, Wohlleben W, et al. Streptocollin, a Type IV Lanthipeptide Produced by *Streptomyces collinus* Tü 365. *ChemBioChem* [Internet]. 2015 Dec 1;16(18):2615–23. Available from: <https://doi.org/10.1002/cbic.201500377>
 297. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* [Internet]. 2010 Mar 8;11:119. Available from: <https://pubmed.ncbi.nlm.nih.gov/20211023>
 298. Halary S, McInerney JO, Lopez P, Baptiste E. EGN: A wizard for construction of gene and genome similarity networks. *BMC Evol Biol*. 2013;13(1):9.
 299. Zaburanyi N, Rabyk M, Ostash B, Fedorenko V, Luzhetskyy A. Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genomics*. 2014 Feb;15(1):97.

300. Kallifidas D, Jiang G, Ding Y, Luesch H. Rational engineering of *Streptomyces albus* J1074 for the overexpression of secondary metabolite gene clusters. *Microb Cell Fact* [Internet]. 2018 Feb 17;17(1):25. Available from: <https://pubmed.ncbi.nlm.nih.gov/29454348>
301. Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: Protein homology by domain architecture. *Genome Res*. 2002 Oct;12(10):1619–23.
302. Zallot R, Oberg N, Gerlt JA. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* [Internet]. 2019 Oct 15;58(41):4169–82. Available from: <https://doi.org/10.1021/acs.biochem.9b00735>
303. Kloosterman AM, Cimermancic P, Elsayed SS, Du C, Hadjithomas M, Donia MS, et al. Expansion of RiPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lanthipeptides. *PLOS Biol* [Internet]. 2020 Dec 22;18(12):e3001026. Available from: <https://doi.org/10.1371/journal.pbio.3001026>
304. Cox CL, Doroghazi JR, Mitchell DA. The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles. *BMC Genomics*. 2015;16(1):16.
305. Miller MA, Pfeiffer W, Schwartz T. The CIPRES Science Gateway: A Community Resource for Phylogenetic Analyses. In: *Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery* [Internet]. New York, NY, USA: Association for Computing Machinery; 2011. (TG '11). Available from: <https://doi.org/10.1145/2016741.2016785>
306. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res* [Internet]. 2009 May 20;37(SUPPL. 2):W202–8. Available from: <https://doi.org/10.1093/nar/gkp335>
307. Boonkerd N. Symbiotic association between Frankia and actinorhizal plants. In: Malik KA, Mirza MS, Ladha JK, editors. *Nitrogen Fixation with Non-Legumes* [Internet]. Dordrecht: Springer Netherlands; 1998. p. 327–31. Available from: https://doi.org/10.1007/978-94-011-5232-7_38
308. Medema MH, Takano E, Breitling R. Detecting sequence homology at the gene cluster level with multigeneblast. *Mol Biol Evol* [Internet]. 2013/02/14. 2013 May;30(5):1218–23. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23412913>
309. Gust B, Challis GL, Fowler K, Kieser T, Chater KF. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the

- sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A*. 2003 Feb;100(4):1541–6.
310. Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, et al. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol*. 2012 May;30(5):440–6.
 311. Mosberg JA, Lajoie MJ, Church GM. Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. *Genetics* [Internet]. 2010/09/02. 2010 Nov;186(3):791–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/20813883>
 312. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* [Internet]. 2000 Jun 6;97(12):6640–5. Available from: <http://www.pnas.org/content/97/12/6640.abstract>
 313. Bellaiche Y, Mogila V, Perrimon N. I-SceI endonuclease, a new tool for studying DNA double strand break repair mechanisms in *Drosophila*. *Genetics* [Internet]. 1999 Jul;152(3):1037–44. Available from: <https://pubmed.ncbi.nlm.nih.gov/10388822>
 314. Carroll D. Genome engineering with zinc-finger nucleases. *Genetics* [Internet]. 2011 Aug;188(4):773–82. Available from: <https://pubmed.ncbi.nlm.nih.gov/21828278>
 315. Joung JK, Sander JD. TALENs: A widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* [Internet]. 2012/11/21. 2013 Jan;14(1):49–55. Available from: <https://pubmed.ncbi.nlm.nih.gov/23169466>
 316. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* (80-). 2012 Aug;337(6096):816–21.
 317. Cobb RE, Wang Y, Zhao H. High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synth Biol*. 2015 Jun;4(6):723–8.
 318. Enquist LW, Kikuchi A, Weisberg RA. The role of λ integrase in integration and excision. *Cold Spring Harb Symp Quant Biol*. 1979;43(2):1115–20.
 319. Kuhstoss S, Rao RN. Analysis of the integration function of the streptomyces bacteriophage ϕ C31. *J Mol Biol* [Internet]. 1991;222(4):897–908. Available from: <http://www.sciencedirect.com/science/article/pii/002228369190584S>
 320. Gregory MA, Till R, Smith MCM. Integration site for *Streptomyces* phage ϕ BT1 and development of site-specific integrating vectors. *J Bacteriol* [Internet]. 2003 Sep;185(17):5320–3. Available from: <https://pubmed.ncbi.nlm.nih.gov/12923110>
 321. Jensen PR, Chavarria KL, Fenical W, Moore BS, Ziemert N. Challenges and

- triumphs to genomics-based natural product discovery. *J Ind Microbiol Biotechnol* [Internet]. 2013/10/09. 2014 Feb;41(2):203–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/24104399>
322. Greunke C, Duell ER, D'Agostino PM, Glöckle A, Lamm K, Gulder TAM. Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab Eng*. 2018 May;47:334–45.
 323. Kouprina N, Larionov V. TAR Cloning: Perspectives for Functional Genomics, Biomedicine, and Biotechnology. *Mol Ther - Methods Clin Dev* [Internet]. 2019;14:16–26. Available from: <http://www.sciencedirect.com/science/article/pii/S2329050119300506>
 324. Larionov V, Kouprina N, Graves J, Chen XN, Korenberg JR, Resnick MA. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc Natl Acad Sci U S A* [Internet]. 1996 Jan 9;93(1):491–6. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/8552668>
 325. Larionov V, Kouprina N, Solomon G, Barrett JC, Resnick MA. Direct isolation of human *brca2* gene by transformation-associated recombination in yeast. *Proc Natl Acad Sci U S A*. 1997 Jul;94(14):7384–7.
 326. Noskov VN, Kouprina N, Leem SH, Ouspenski I, Barrett C, Larionov V. A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. *BMC Genomics*. 2003 Apr;4(1):16.
 327. Jiang W, Zhao X, Gabrieli T, Lou C, Ebenstein Y, Zhu TF. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun* [Internet]. 2015;6(1):8101. Available from: <https://doi.org/10.1038/ncomms9101>
 328. Labeda DP, Doroghazi JR, Ju KS, Metcalf WW. Taxonomic evaluation of *Streptomyces albus* and related species using multilocus sequence analysis and proposals to emend the description of *Streptomyces albus* and describe *Streptomyces pathocidini* sp. nov. *Int J Syst Evol Microbiol* [Internet]. 2013/11/25. 2014 Mar;64(PART 3):894–900. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/24277863>
 329. Myronovskyi M, Rosenkränzer B, Nadmid S, Pujic P, Normand P, Luzhetskyy A. Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab Eng*. 2018 Sep;49:316–24.
 330. Yorgey P, Lee J, Kördel J, Vivas E, Warner P, Jebaratnam D, et al. Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. *Proc Natl Acad Sci U S A* [Internet]. 1994 May 10;91(10):4519–

23. Available from: <http://www.pnas.org/content/91/10/4519.abstract>
331. Eyles TH, Vior NM, Truman AW. Rapid and Robust Yeast-Mediated Pathway Refactoring Generates Multiple New Bottromycin-Related Metabolites. *ACS Synth Biol* [Internet]. 2018 May 18;7(5):1211–8. Available from: <https://doi.org/10.1021/acssynbio.8b00038>
332. Tang X, Li J, Millán-Aguíñaga N, Zhang JJ, O'Neill EC, Ugalde JA, et al. Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chem Biol*. 2015;10(12):2841–9.
333. Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: A model system for the study of recombination. *Proc Natl Acad Sci U S A* [Internet]. 1981 Oct 1;78(10):6354–8. Available from: <http://www.pnas.org/content/78/10/6354.abstract>
334. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* [Internet]. 2009;6(5):343–5. Available from: <https://doi.org/10.1038/nmeth.1318>
335. Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, et al. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci U S A* [Internet]. 2014;111(5):1957–62. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3918841/pdf/pnas.201319584.pdf>
336. Gomez-Escribano JP, Bibb MJ. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol* [Internet]. 2011 Mar 1;4(2):207–15. Available from: <https://doi.org/10.1111/j.1751-7915.2010.00219.x>
337. Chater KF, Wilde LC. *Streptomyces albus* G mutants defective in the SalGI restriction-modification system. *J Gen Microbiol* [Internet]. 1980;116(2):323–34. Available from: <https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-116-2-323>
338. Kallifidas D, Kang HS, Brady SF. Tetarimycin A, an MRSA-active antibiotic identified through induced expression of environmental DNA gene clusters. *J Am Chem Soc* [Internet]. 2012 Dec 5;134(48):19552–5. Available from: <https://doi.org/10.1021/ja3093828>
339. Lombó F, Velasco A, Castro A, De La Calle F, Braña AF, Sánchez-Puelles JM, et al. Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two *Streptomyces* species. *ChemBioChem* [Internet]. 2006 Feb 6;7(2):366–76. Available from:

<https://doi.org/10.1002/cbic.200500325>

340. Bilyk O, Sekurova ON, Zotchev SB, Luzhetskyy A. Cloning and heterologous expression of the greccocycline biosynthetic gene cluster. PLoS One [Internet]. 2016 Jul 13;11(7):e0158682. Available from: <https://doi.org/10.1371/journal.pone.0158682>
341. Feng Z, Kallifidas D, Brady SF. Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. Proc Natl Acad Sci U S A [Internet]. 2011 Aug 2;108(31):12629–34. Available from: <http://www.pnas.org/content/108/31/12629.abstract>
342. Eckert M. Max von Laue and the discovery of X-ray diffraction in 1912. Ann Phys [Internet]. 2012 May 3;524(5):A83–5. Available from: <https://doi.org/10.1002/andp.201200724>
343. HODGKIN DC. The X-ray analysis of the structure of penicillin. Adv Sci. 1949 Jul;6(22):85–9.
344. Hodgkin DC, Kamper J, MacKay M, Pickworth J, Trueblood KN, White JG. Structure of vitamin B12. Nature [Internet]. 1956;178(4524):64–6. Available from: <https://doi.org/10.1038/178064a0>
345. Blundell T, Dodson G, Hodgkin D, Mercola D. Insulin: The structure in the crystal and its reflection in chemistry and biology. In: Anfinsen CB, Edsall JT, Richards FMBT-A in PC, editors. Advances in Protein Chemistry [Internet]. Academic Press; 1972. p. 279–86. Available from: <http://www.sciencedirect.com/science/article/pii/S0065323308601436>
346. Rabi II, Zacharias JR, Millman S, Kusch P. A new method of measuring nuclear magnetic moment [5]. Phys Rev [Internet]. 1938 Feb 15;53(4):318. Available from: <https://link.aps.org/doi/10.1103/PhysRev.53.318>
347. Aue WP, Bartholdi E, Ernst RR. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. J Chem Phys [Internet]. 1976 Mar 1;64(5):2229–46. Available from: <https://doi.org/10.1063/1.432450>
348. Jones CG, Martynowycz MW, Hattne J, Fulton TJ, Stoltz BM, Rodriguez JA, et al. The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. ACS Cent Sci [Internet]. 2018 Nov 28;4(11):1587–92. Available from: <https://doi.org/10.1021/acscentsci.8b00760>
349. Shi D, Nannenga BL, Iadanza MG, Gonen T. Three-dimensional electron crystallography of protein microcrystals. Harrison SC, editor. Elife [Internet]. 2013;2:e01345. Available from: <https://doi.org/10.7554/eLife.01345>
350. Heatley NG. A method for the assay of penicillin. Biochem J [Internet]. 1944;38(1):61–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/16747749>
351. Andrews JM. Determination of minimum inhibitory concentrations. J Antimicrob

- Chemother. 2001 Jul;48(SUPPL. 1):5–16.
352. Bueno C, Villegas ML, Bertolotti SG, Previtali CM, Neumann MG, Encinas M V. The Excited-State Interaction of Resazurin and Resorufin with Amines in Aqueous Solutions. *Photophysics and Photochemical Reaction*. *Photochem Photobiol* [Internet]. 2002 Oct 1;76(4):385–90. Available from: [https://doi.org/10.1562/0031-8655\(2002\)0760385TESIOR2.0.CO2](https://doi.org/10.1562/0031-8655(2002)0760385TESIOR2.0.CO2)
 353. Lynne AM, Haarmann D, Loudon BC. Use of Blue Agar CAS Assay for Siderophore Detection. *J Microbiol Biol Educ* [Internet]. 2011 May 19;12(1):51–3. Available from: <https://pubmed.ncbi.nlm.nih.gov/23653742>
 354. Mehnert M, Retamal-Morales G, Schwabe R, Vater S, Heine T, Gloria LJ, et al. Revisiting the chrome azurol S assay for various metal ions. *Solid State Phenom* [Internet]. 2017;262 SSP:509–12. Available from: <https://www.scientific.net/SSP.262.509>
 355. Bandara HMD, Kennedy DP, Akin E, Incarvito CD, Burdette SC. Photoinduced release of Zn²⁺ with ZinCleave-1: A nitrobenzyl-based caged complex. *Inorg Chem* [Internet]. 2009 Sep 7;48(17):8445–55. Available from: <https://doi.org/10.1021/ic901062n>
 356. Prabhu V, Chatson B, Abrams G, King J. ¹³C chemical shifts of 20 free amino acids and their use in detection by NMR of free amino acids in intact plants of *Arabidopsis*. *J Plant Physiol* [Internet]. 1996;149(3–4):246–50. Available from: <http://www.sciencedirect.com/science/article/pii/S0176161796801167>
 357. Kaneda M. Studies on bottromycins I. ¹H and ¹³C NMR assignments of bottromycin A₂, the main component of the complex. *J Antibiot (Tokyo)*. 1992 May;45(5):792–6.
 358. Schwalen CJ, Hudson GA, Kosol S, Mahanta N, Challis GL, Mitchell DA. In Vitro Biosynthetic Studies of Bottromycin Expand the Enzymatic Capabilities of the YcaO Superfamily. *J Am Chem Soc*. 2017;139(50):18154–7.
 359. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* [Internet]. 2014 Apr 21;42(W1):W320–4. Available from: <https://doi.org/10.1093/nar/gku316>
 360. Ghilarov D, Stevenson CEM, Travin DY, Piskunova J, Serebryakova M, Maxwell A, et al. Architecture of Microcin B17 Synthetase: An Octameric Protein Complex Converting a Ribosomally Synthesized Peptide into a DNA Gyrase Poison. *Mol Cell* [Internet]. 2019;73(4):749–762.e5. Available from: <http://www.sciencedirect.com/science/article/pii/S1097276518310025>
 361. Wang M, Jarmusch AK, Vargas F, Aksenov AA, Gauglitz JM, Weldon K, et al. Mass spectrometry searches using MASST. *Nat Biotechnol* [Internet]. 2020;38(1):23–6. Available from: <https://doi.org/10.1038/s41587-019-0375-9>

362. Button JE, Silhavy TJ, Ruiz N. A suppressor of cell death caused by the loss of σ^E downregulates extracytoplasmic stress responses and outer membrane vesicle production in *Escherichia coli*. *J Bacteriol* [Internet]. 2007 Mar 1;189(5):1523–30. Available from: <http://jb.asm.org/content/189/5/1523.abstract>
363. Tong Y, Whitford CM, Robertsen HL, Blin K, Jørgensen TS, Klitgaard AK, et al. Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proc Natl Acad Sci U S A* [Internet]. 2019 Oct 8;116(41):20366–75. Available from: <http://www.pnas.org/content/116/41/20366.abstract>
364. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* [Internet]. 2013;8(11):2180–96. Available from: <https://doi.org/10.1038/nprot.2013.132>
365. Young TS, Dorrestein PC, Walsh CT. Codon randomization for rapid exploration of chemical space in thiopeptide antibiotic variants. *Chem Biol*. 2012 Dec;19(12):1600–10.
366. Ruffner DE, Schmidt EW, Heemstra JR. Assessing the combinatorial potential of the RiPP cyanobactin tru pathway. *ACS Synth Biol* [Internet]. 2015;4(4):482–92. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4410914/pdf/sb500267d.pdf>
367. Yang X, Lennard KR, He C, Walker MC, Ball AT, Doigneaux C, et al. A lanthipeptide library used to identify a protein-protein interaction inhibitor article. *Nat Chem Biol* [Internet]. 2018;14(4):375–80. Available from: <https://doi.org/10.1038/s41589-018-0008-5>
368. Kumamoto T. Amidines and Guanidines in Natural Products and Medicines [Internet]. *Superbases for Organic Synthesis: Guanidines, Amidines, Phosphazenes and Related Organocatalysts*. 2009. p. 295–313. (Wiley Online Books). Available from: <https://doi.org/10.1002/9780470740859.ch10>
369. Jost M, Greie JC, Stemmer N, Wilking SD, Altendorf K, Sewald N. The first total synthesis of efrapeptin C. *Angew Chemie - Int Ed* [Internet]. 2002 Nov 15;41(22):4267–9. Available from: [https://doi.org/10.1002/1521-3773\(20021115\)41:22%3C4267::AID-ANIE4267%3E3.0.CO](https://doi.org/10.1002/1521-3773(20021115)41:22%3C4267::AID-ANIE4267%3E3.0.CO)
370. Bushley KE, Raja R, Jaiswal P, Cumbie JS, Nonogaki M, Boyd AE, et al. The genome of *tolypocladium inflatum*: evolution, organization, and expression of the cyclosporin biosynthetic gene cluster. *PLoS Genet* [Internet]. 2013/06/20. 2013 Jun;9(6):e1003496–e1003496. Available from: <https://pubmed.ncbi.nlm.nih.gov/23818858>
371. Krasnoff SB, Gupta S, Leger RJS, Renwick JAA, Roberts DW. Antifungal and insecticidal properties of the efrapeptins: Metabolites of the fungus

- Tolypocladium niveum. J Invertebr Pathol [Internet]. 1991;58(2):180–8. Available from: <http://www.sciencedirect.com/science/article/pii/002220119190062U>
372. Abrahams JP, Buchanan SK, Van Raaij MJ, Fearnley IM, Leslie AGW, Walker JE. The structure of bovine F1-ATPase complexed with the peptide antibiotic efrapeptin. Proc Natl Acad Sci U S A [Internet]. 1996 Sep 3;93(18):9420–4. Available from: <https://pubmed.ncbi.nlm.nih.gov/8790345>
 373. Fredenhagen A, Molleyres LP, Böhlendorf B, Laue G. Structure determination of neoeprapeptins A to N: Peptides with insecticidal activity produced by the fungus Geotrichum candidum. J Antibiot (Tokyo). 2006 May;59(5):267–80.
 374. Richter AA, Mais CN, Czech L, Geyer K, Hoepfner A, Smits SHJ, et al. Biosynthesis of the Stress-Protectant and Chemical Chaperon Ectoine: Biochemistry of the Transaminase EctB. Front Microbiol. 2019;10(December):1–20.
 375. Kuhlmann AU, Hoffmann T, Bursy J, Jebbar M, Bremer E. Ectoine and hydroxyectoine as protectants against osmotic and cold stress: Uptake through the SigB-controlled betaine-choline-carnitine transporter-type carrier ectt from Virgibacillus pantothenticus. J Bacteriol [Internet]. 2011 Sep 15;193(18):4699–708. Available from: <http://jb.asm.org/content/193/18/4699.abstract>
 376. Aoyama T, Kojima F, Imada C, Muraoka Y, Naganawa H, Okami Y, et al. Pyrostatins a and b, new inhibitors of n-acetyl- β -D-glucosaminidase, produced by streptomyces sp. SA-3501. J Enzyme Inhib Med Chem. 1995;8(4):223–32.
 377. Appel D, Lentzen G. The pyrostatins A and B do not inhibit N-acetyl- β -D-glucosaminidase. J Enzyme Inhib Med Chem. 2009 Oct;24(5):1106–8.
 378. Peters L, Wright AD, Krick A, König GM. Variation of brominated indoles and terpenoids within single and different colonies of the marine bryozoan Flustra foliacea. J Chem Ecol [Internet]. 2004;30(6):1165–81. Available from: <https://doi.org/10.1023/B:JOEC.0000030270.65594.f4>
 379. Lindel T, Bräuchle L, Golz G, Böhler P. Total synthesis of flustramine C via dimethylallyl rearrangement. Org Lett [Internet]. 2007 Jan 1;9(2):283–6. Available from: <https://doi.org/10.1021/ol0627348>
 380. Kouprina N, Annab L, Graves J, Afshari C, Barrett JC, Resnick MA, et al. Functional copies of a human gene can be directly isolated by transformation-associated recombination cloning with a small 3' end target sequence. Proc Natl Acad Sci U S A [Internet]. 1998 Apr 14;95(8):4469–74. Available from: <http://www.pnas.org/content/95/8/4469.abstract>
 381. MacNeil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, MacNeil T. Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. Gene. 1992 Feb;111(1):61–8.

382. Costantino N, Court DL. Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci U S A* [Internet]. 2003 Dec 23;100(26):15748–53. Available from: <http://www.pnas.org/content/100/26/15748.abstract>
383. Bierman M, Logan R, O'Brien K, Seno ET, Nagaraja Rao R, Schonher BE. Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene*. 1992;116(1):43–9.
384. Hong HJ, Hutchings MI, Hill LM, Buttner MJ. The role of the novel fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*. *J Biol Chem*. 2005 Apr;280(13):13055–61.
385. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003 Nov;13(11):2498–504.
386. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res* [Internet]. 2019 Apr 1;47(W1):W256–9. Available from: <https://doi.org/10.1093/nar/gkz239>
387. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016 Jul;33(7):1870–4.
388. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*. 1994;2:28–36.
389. Phyre2 [Internet]. Available from: <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>
390. Alexander DB, Zuberer DA. Use of chrome azurol S reagents to evaluate siderophore production by rhizosphere bacteria. *Biol Fertil Soils* [Internet]. 1991;12(1):39–45. Available from: <https://doi.org/10.1007/BF00369386>

Appendices

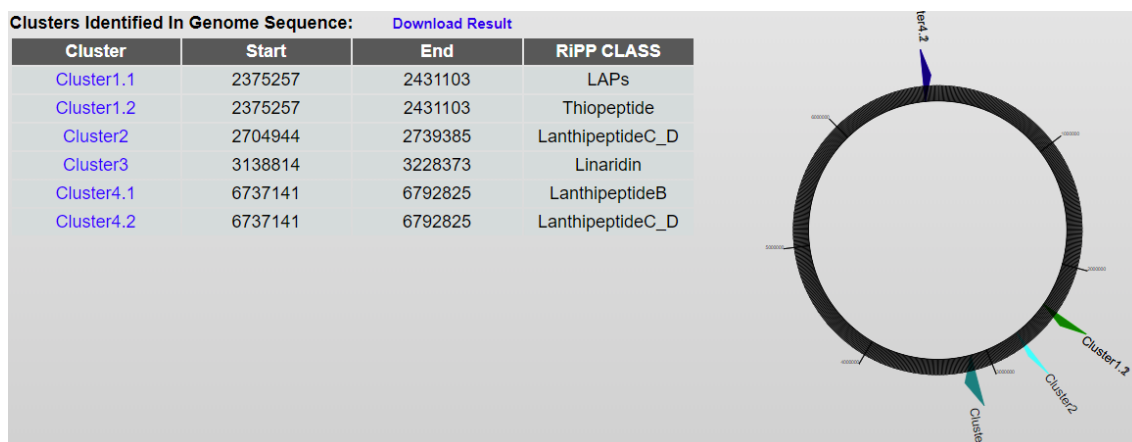


Figure 78: Genome mining output from RiPPMiner analysis of *S. albus* J1074 genome.

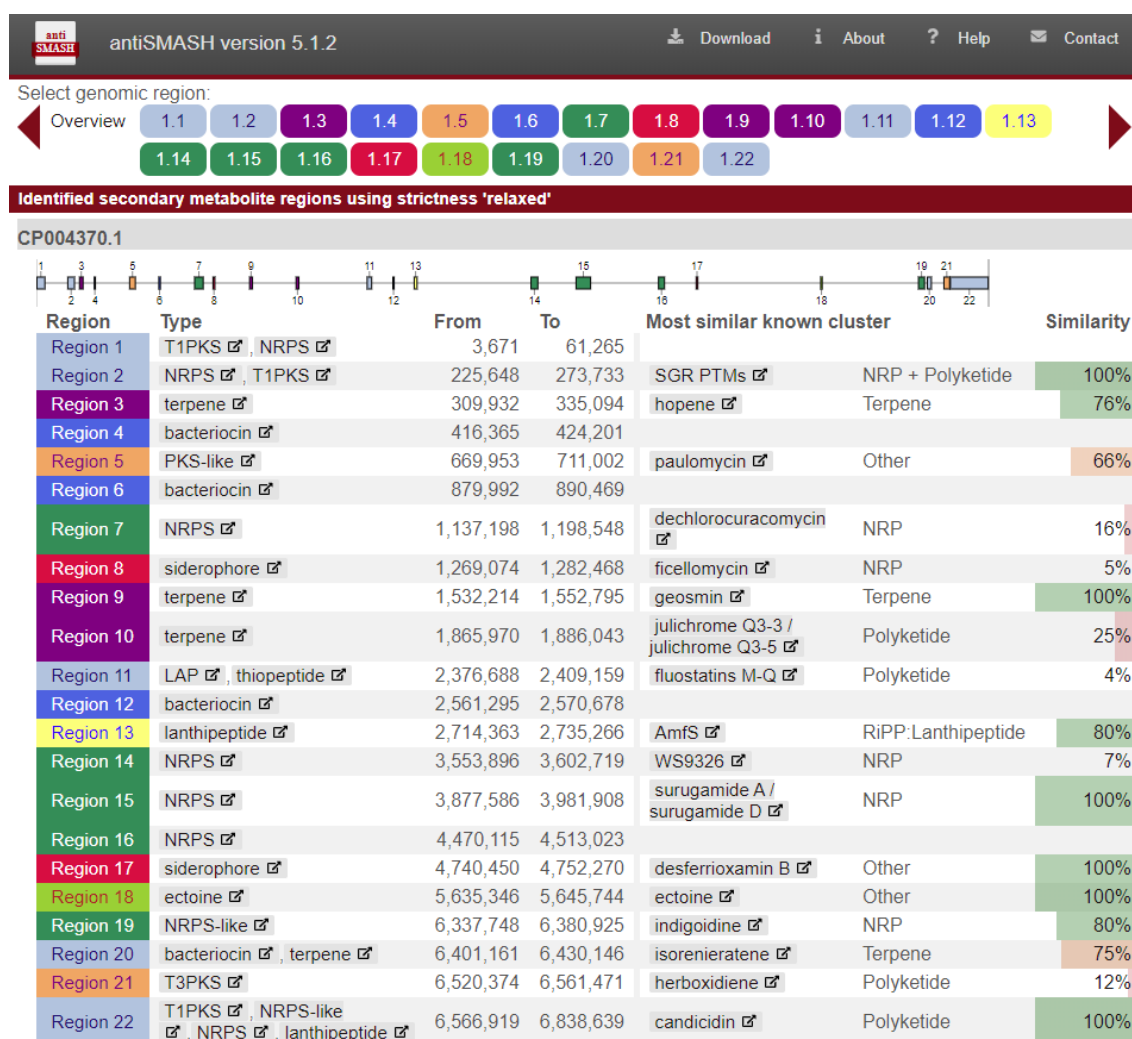


Figure 79: Genome mining output from antiSMASH analysis of *S. albus* J1074 genome.

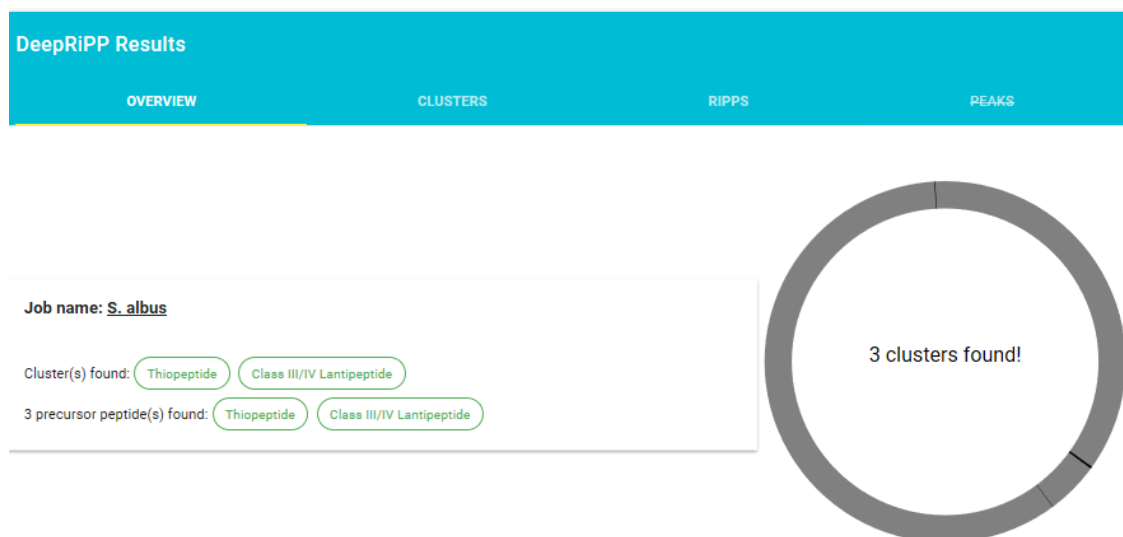


Figure 80: Genome mining output from DeepRiPP analysis of *S. albus* J1074 genome.

Run summary				
Number of files analyzed	2			
Number of DNA fragments analyzed	1			
Total bases in all DNA	6841649			
Number of AOI's A reas O f I nterest)	3			

AOI	start	end	Class	Fasta header
CP0043701.0.AOI_01	2713069	2733195	204.1;SAL	CP004370.1 Streptomyces albidoflavus strain J1074, complete genome
CP0043701.0.AOI_02	6760442	6780442	Lanthipeptide_class_IV	CP004370.1 Streptomyces albidoflavus strain J1074, complete genome
CP0043701.0.AOI_03	6396199	6416961	82.3;Linocin-M18	CP004370.1 Streptomyces albidoflavus strain J1074, complete genome

Figure 81: Genome mining output from BAGEL analysis of *S. albus* J1074 genome.

Clusters

Sequence: S._albus_genome.fasta.json

Search configuration: [download](#)

JSON report: [download](#)

These results will be available at </prism/tasks/20200327-1731-403998930/> for 60 days.

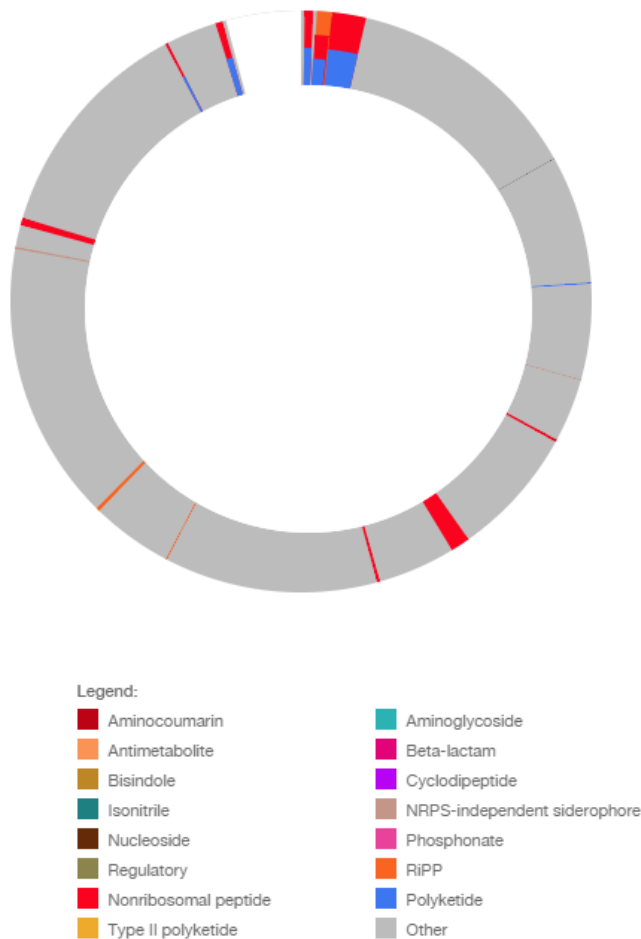
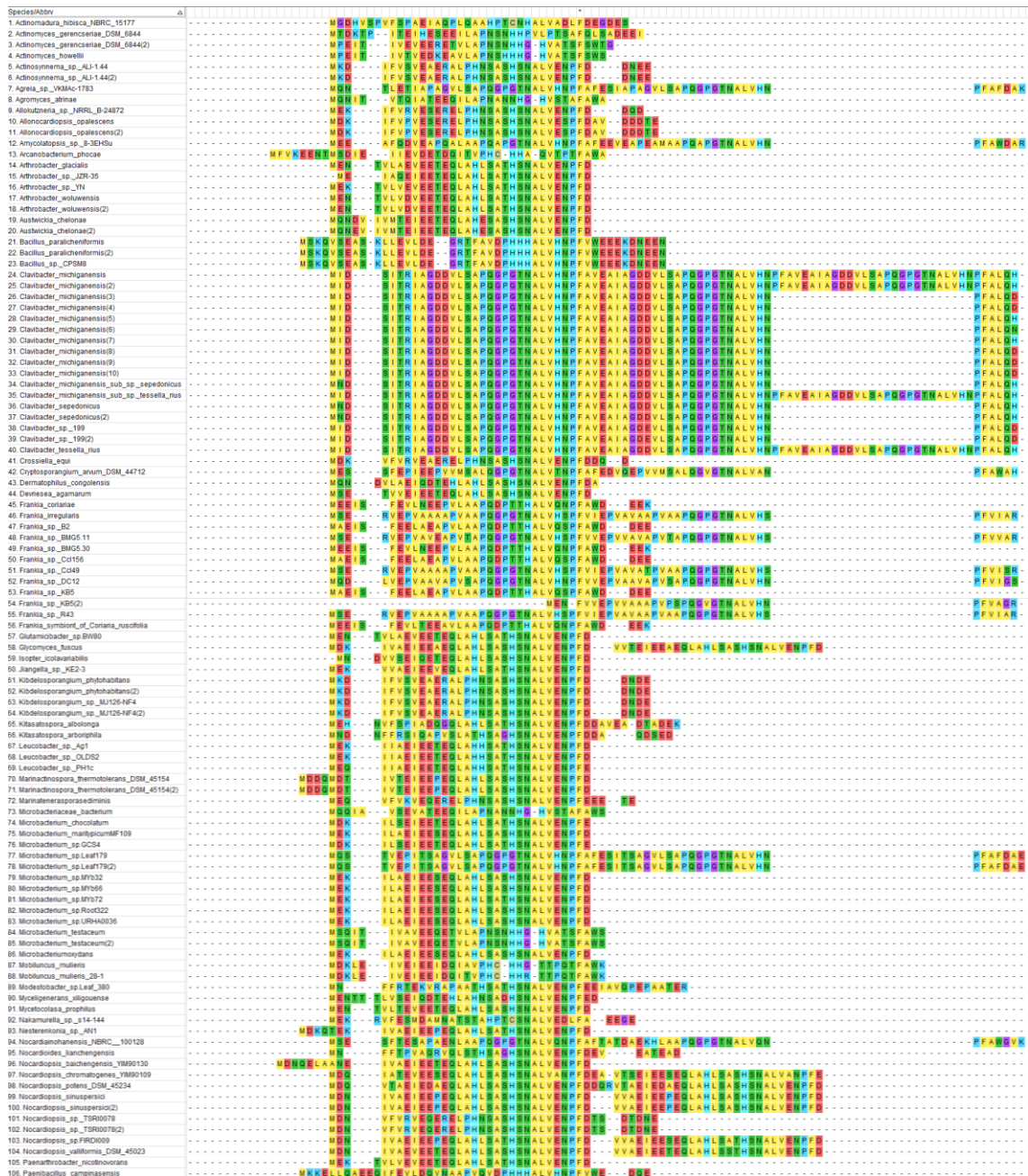


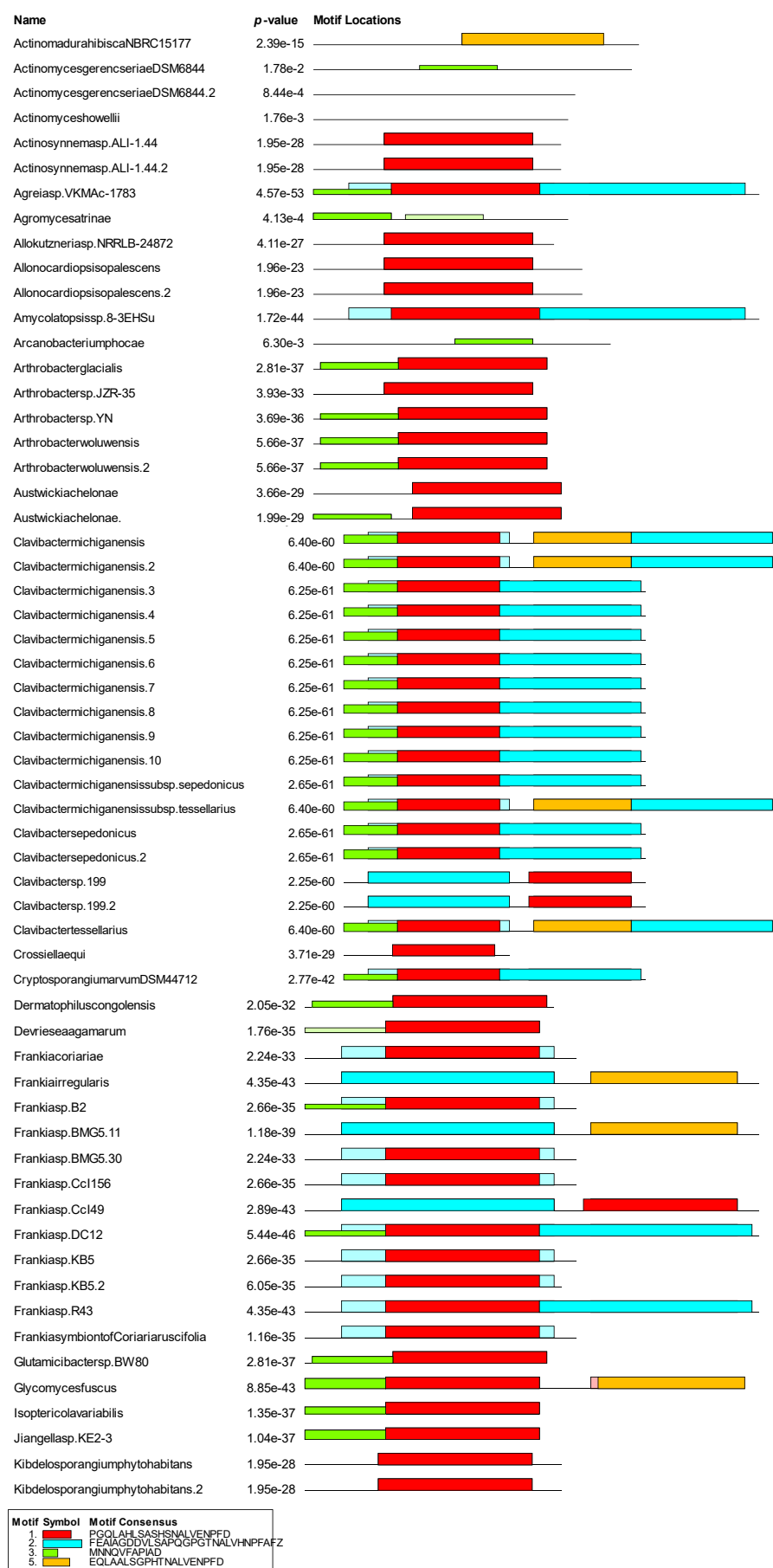
Figure 82: Genome mining output from PRISM analysis of *S. albus* J1074 genome.

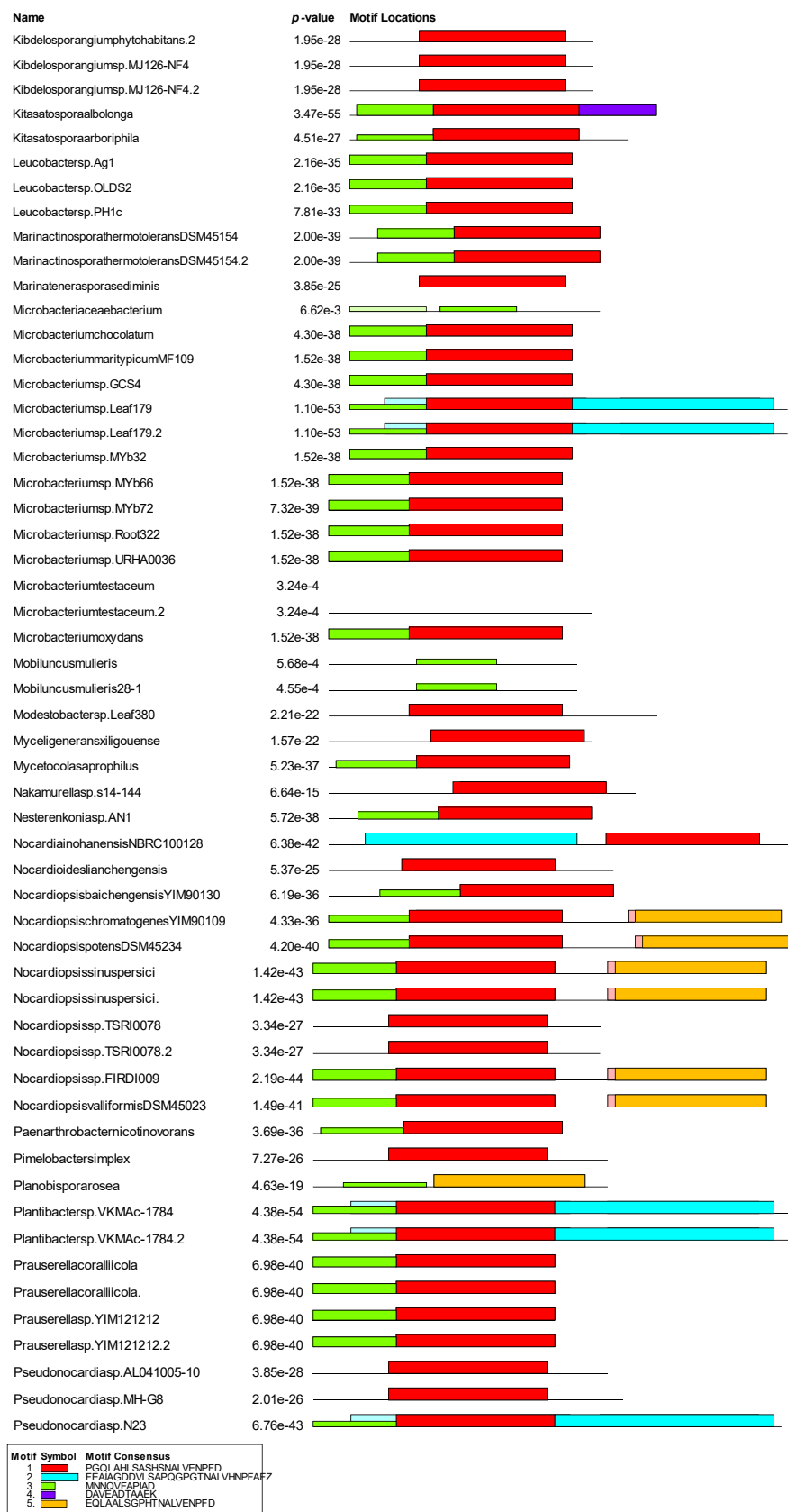
Figure 83 (below): Alignment of streptomidine-like precursor peptides associated with precursor network 1, obtained from RiPPER analysis of *S. albus* J1074 YcaO-domain protein

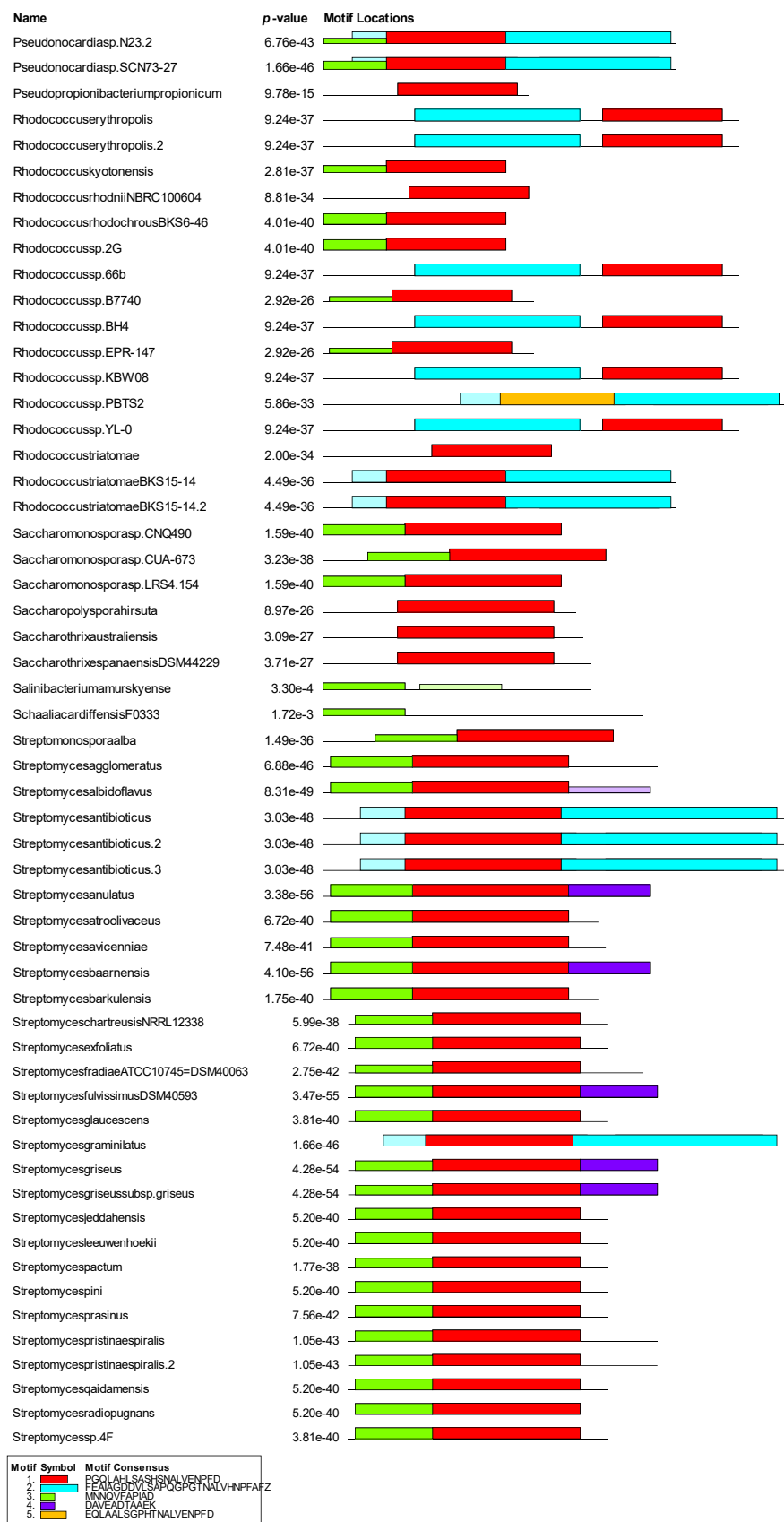


201

Figure 84 (below): MEME output from analysis of 231 streptomidine-like precursor peptide sequences







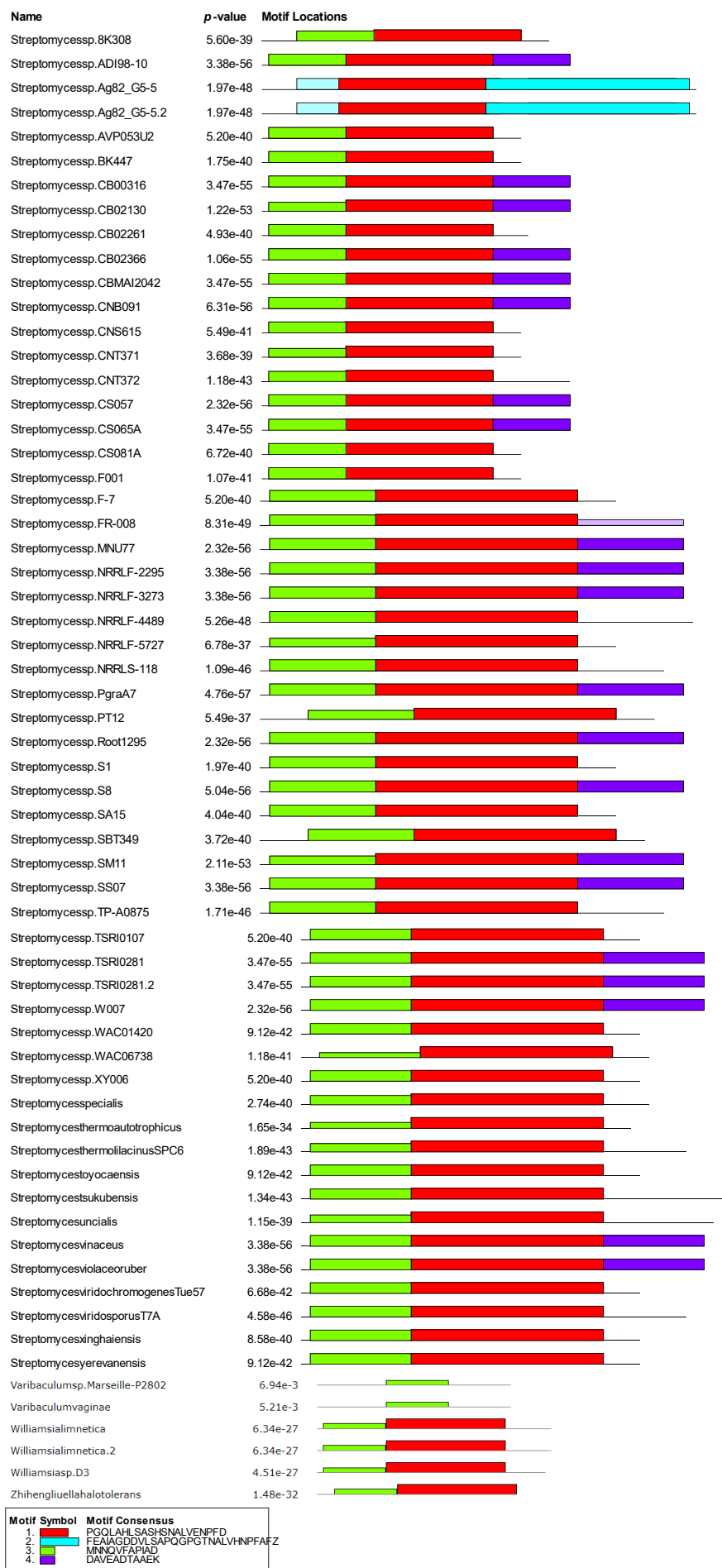
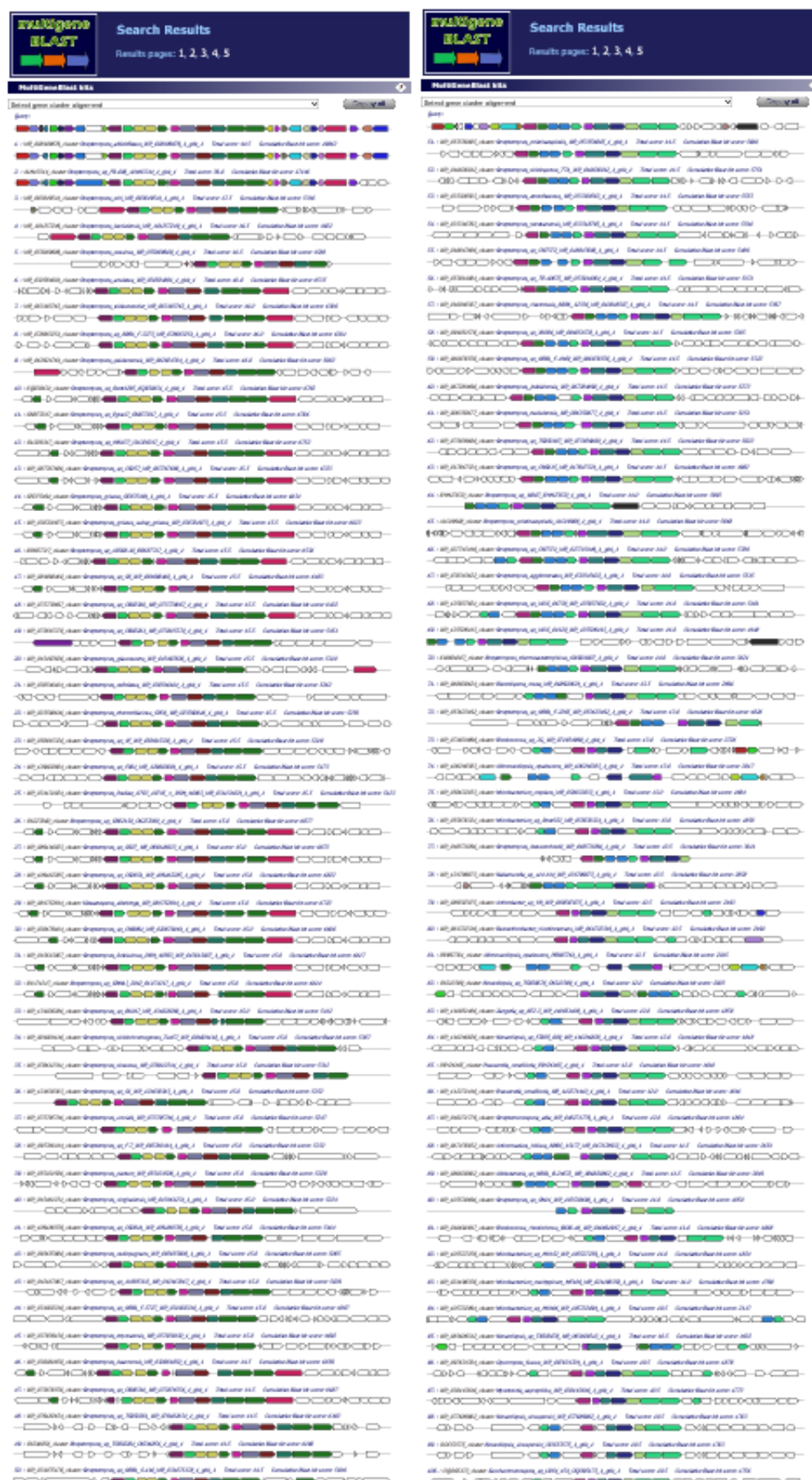


Figure 85 (below): MultiGeneBlast output from analysis of *Streptomyces albus* (albidoflavus) J1074 biosynthetic gene cluster



[illegible]

Search Results

Results pages: 1, 2, 3, 4, 5

RefSeqGene Build 104

Selected gene cluster: *algebra* and *gene*

100 RefSeqGene Build 104

1. ALG1, 2. ALG2, 3. ALG3, 4. ALG4, 5. ALG5, 6. ALG6, 7. ALG7, 8. ALG8, 9. ALG9, 10. ALG10, 11. ALG11, 12. ALG12, 13. ALG13, 14. ALG14, 15. ALG15, 16. ALG16, 17. ALG17, 18. ALG18, 19. ALG19, 20. ALG20, 21. ALG21, 22. ALG22, 23. ALG23, 24. ALG24, 25. ALG25, 26. ALG26, 27. ALG27, 28. ALG28, 29. ALG29, 30. ALG30, 31. ALG31, 32. ALG32, 33. ALG33, 34. ALG34, 35. ALG35, 36. ALG36, 37. ALG37, 38. ALG38, 39. ALG39, 40. ALG40, 41. ALG41, 42. ALG42, 43. ALG43, 44. ALG44, 45. ALG45, 46. ALG46, 47. ALG47, 48. ALG48, 49. ALG49, 50. ALG50, 51. ALG51, 52. ALG52, 53. ALG53, 54. ALG54, 55. ALG55, 56. ALG56, 57. ALG57, 58. ALG58, 59. ALG59, 60. ALG60, 61. ALG61, 62. ALG62, 63. ALG63, 64. ALG64, 65. ALG65, 66. ALG66, 67. ALG67, 68. ALG68, 69. ALG69, 70. ALG70, 71. ALG71, 72. ALG72, 73. ALG73, 74. ALG74, 75. ALG75, 76. ALG76, 77. ALG77, 78. ALG78, 79. ALG79, 80. ALG80, 81. ALG81, 82. ALG82, 83. ALG83, 84. ALG84, 85. ALG85, 86. ALG86, 87. ALG87, 88. ALG88, 89. ALG89, 90. ALG90, 91. ALG91, 92. ALG92, 93. ALG93, 94. ALG94, 95. ALG95, 96. ALG96, 97. ALG97, 98. ALG98, 99. ALG99, 100. ALG100.

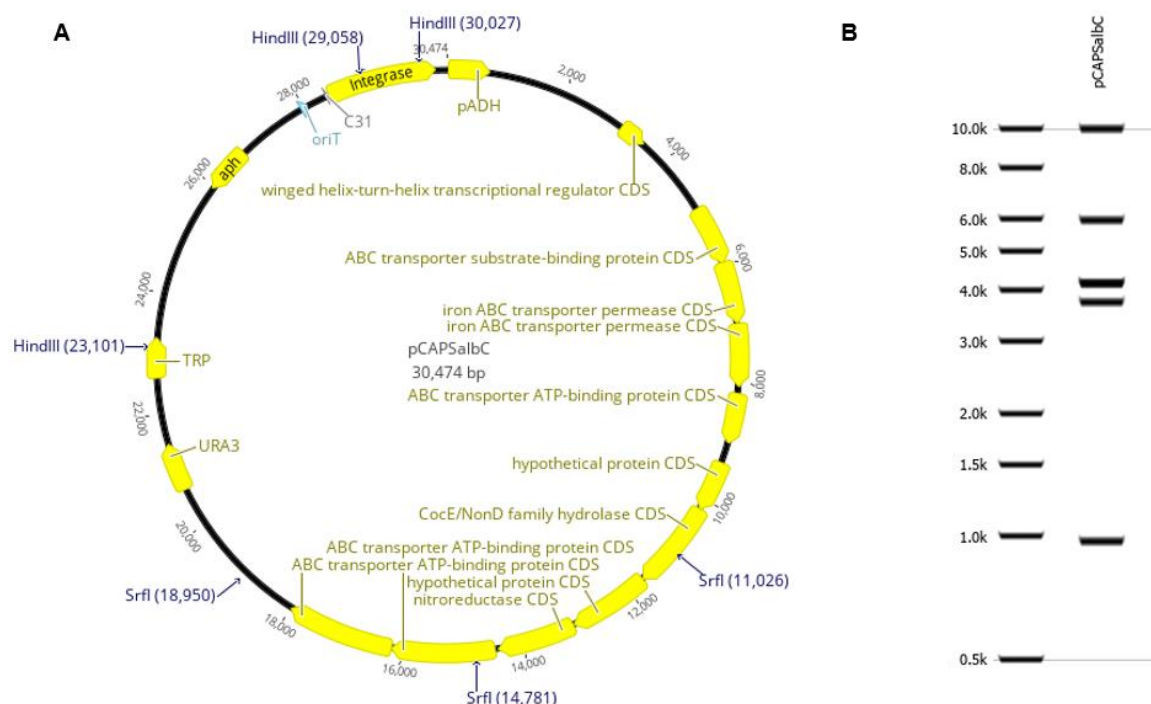


Figure 86: (A) Vector map of pCAPSalbC with the HindIII and SrfI restriction sites highlighted. (B) 'Virtual gel' image of expected DNA bands obtained from an analytical restriction digest of pCAPSalbC with HindIII and SrfI enzymes, obtained from geneious software (263). This is similar to the agarose gel displayed in Figure 31.

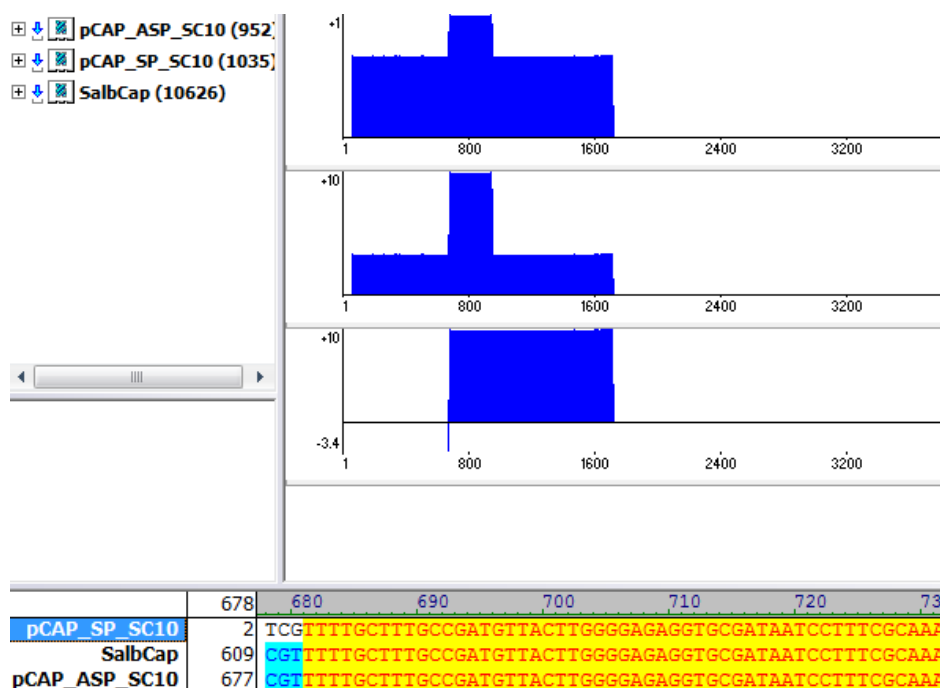


Figure 87: Screenshot of DNA alignment obtained from sequencing of pSalbCAP construct. The expected vector sequence matches the sequences obtained from forward and reverse screening of the plasmid obtained following Gibson assembly of the capture vector.

Table 22: Selected metabolomic data showing metabolites produced by *S. coelicolor* M1146-SalbC compared with the precursor peptide mutant (SalbCΔPP) and medium only (SM12). Analysis carried out using Profiling Solution (Shimadzu) where numbers reflect peak areas for specific metabolites.

Ion m/z	510.2674	324.17	272.1603	314.1695	647.33	409.2343
Ion RT	1.09	0.869	2.423	2.48	0.989	1.302
<i>S. coelicolor</i> M1146:pCAPSalbC1	2499654	2207472	996032	873422	298441	260134
<i>S. coelicolor</i> M1146:pCAPSalbC2	1922639	2825998	846665	962465	268447	281377
<i>S. coelicolor</i> M1146:pCAPSalbC3	2372999	2730479	990778	1160814	307189	214862
<i>S. coelicolor</i> M1146:pCAPSalbCΔPP1	0	0	0	0	0	0
<i>S. coelicolor</i> M1146:pCAPSalbCΔPP2	0	0	0	0	0	0
<i>S. coelicolor</i> M1146:pCAPSalbCΔPP3	0	0	0	0	0	0
SM12	0	0	0	0	0	0

Table 23: Selected metabolomic data showing metabolites produced by *S. coelicolor* M1146-SalbC compared with pathway mutants (ΔamiB, ΔamiC, ΔamiD, ΔamiE, ΔamiX) and medium only (SM12). Analysis carried out as described for Table 21.

Ion m/z	324.1659	510.2698	314.1696	647.3248	332.1785	272.1582	354.1602	510.2688
Ion RT	1.398	1.55	3.025	1.392	3.023	2.983	3.034	1.202
SalbC1	2262279	1475999	1042037	544662	393459	387710	280038	205147
SalbC2	2214873	1171449	847530	545338	329963	410005	252484	197976
ΔamiA1	0	0	0	0	0	0	0	0
ΔamiA2	0	0	0	0	0	0	0	0
ΔamiA3	0	0	0	0	0	0	0	0
ΔamiB1	0	0	0	0	0	0	0	0
ΔamiB2	0	0	0	0	0	0	0	0
ΔamiB3	0	0	0	0	0	0	0	0
ΔamiC1	0	0	0	0	0	0	0	0
ΔamiC2	0	0	0	0	0	0	0	0
ΔamiC3	0	0	0	0	0	0	0	0
ΔamiD1	0	0	0	0	0	0	0	0
ΔamiD2	0	0	0	0	0	0	0	0
ΔamiD3	0	0	0	0	0	0	0	0
ΔamiE1	0	0	1605224	0	887866	796578	443885	0
ΔamiE2	0	0	2048790	0	1116939	1351619	483075	0
ΔamiE3	0	0	1588020	0	712023	943842	389179	0
ΔamiX1	1057241	810404	2825272	290000	1980880	1570774	649966	172142
ΔamiX2	1172300	763166	2569195	237703	1813192	1699116	604358	149342
ΔamiX3	1139022	660021	2451654	223171	1386092	1560029	585720	169857
SM12 1	0	0	0	0	0	0	0	0

Table 24: Selected metabolomic data showing metabolites produced by *S. coelicolor* M1146-SalbC compared with pathway mutants: iron transporter deletion (Δ amiF1-4), oxygenase deletion (Δ Oxy), peptide methionine sulfoxide reductase deletion (Δ PM), MarR deletion (Δ MarR), ABC transporter deletion (Δ T1-2) and acetyltransferase deletion (Δ AT). Analysis carried out as described for Table 21.

Ion m/z	324.1652	409.2112	510.2665
Ion RT	1.335	1.393	1.503
SalbC1	2262502	304396	104359
SalbC2	2715880	547614	294043
SalbC3	2252375	234775	110043
Δ amiF1-4 1	0	0	0
Δ amiF1-4 2	0	0	0
Δ amiF1-4 3	0	0	0
Δ amiT1-2 1	389484	152524	0
Δ amiT1-2 2	206483	130887	0
Δ amiT1-2 3	0	0	0
Δ PM1	3046506	497514	158387
Δ PM2	2325230	368514	126083
Δ PM3	2642271	313950	131314
Δ Oxy1	2638215	170407	167448
Δ Oxy2	3555376	183279	222399
Δ Oxy3	141632	0	0
Δ MarR1	3931198	929006	1113021
Δ MarR2	2664038	446943	268230
Δ MarR3	4194032	711481	524694
Δ AT1	2608474	174412	300238
Δ AT2	129689	0	0
Δ AT3	0	0	0

Table 25: Selected metabolomic data showing metabolites produced by *S. albus* J1074 wild type compared with the pathway-disrupted mutant (*S. albus* YH mutant) and medium only (SM12). Analysis carried out as described for Table 21.

Ion m/z	324.1662	647.3265	510.2657	409.2192	272.1584	314.1664
Ion RT	1.755	1.738	2.521	0.8	3.978	4.044
<i>S. albus</i> J1074 WT1	2971974	544327	309829	0	316550	0
<i>S. albus</i> J1074 WT2	3324012	573868	307114	300468	249284	0
<i>S. albus</i> J1074 WT3	3246270	484952	259442	252018	297315	0
<i>S. albus</i> YH mutant1	0	0	0	0	1822945	265587
<i>S. albus</i> YH mutant2	0	0	0	0	0	0
<i>S. albus</i> YH mutant3	0	0	0	0	0	315791
SM12	0	0	0	0	0	0

1788	HLS	-18.0106,15.9949,2.0157,-18.0106	+H	338.182304
1789	HLS	2.0157,-18.0106,-2.0157	+H	338.182304
1790	HLS	-2.0157,2.0157,-18.0106	+H	338.182304
1791	HLS	-2.0157,-18.0106,2.0157	+H	338.182304
1792	HLS	-18.0106	+H	338.182304
1793	HLS	-18.0106,2.0157,-2.0157	+H	338.182304
3397	HLSA	-18.0106,15.9949,2.0157,-18.0106	+H	409.219418
3398	HLSA	2.0157,-18.0106,-2.0157	+H	409.219418
3399	HLSA	-2.0157,2.0157,-18.0106	+H	409.219418
3400	HLSA	-2.0157,-18.0106,2.0157	+H	409.219418
3401	HLSA	-18.0106	+H	409.219418
3402	HLSA	-18.0106,2.0157,-2.0157	+H	409.219418
3403	AHLS	-18.0106,15.9949,2.0157,-18.0106	+H	409.219418
3404	AHLS	2.0157,-18.0106,-2.0157	+H	409.219418
3405	AHLS	-2.0157,2.0157,-18.0106	+H	409.219418
3406	AHLS	-2.0157,-18.0106,2.0157	+H	409.219418
3407	AHLS	-18.0106	+H	409.219418
3408	AHLS	-18.0106,2.0157,-2.0157	+H	409.219418
5772	HLSAT	-2.0157,-18.0106,2.0157	+H	510.267097
5773	LSATH	-2.0157,-18.0106,2.0157	+H	510.267097
5774	HLSAT	-2.0157,2.0157,-18.0106	+H	510.267097
5775	HLSAT	-18.0106,15.9949,2.0157,-18.0106	+H	510.267097
5776	HLSAT	-18.0106	+H	510.267097
5777	HLSAT	-18.0106,2.0157,-2.0157	+H	510.267097
5778	LSATH	-2.0157,2.0157,-18.0106	+H	510.267097
5779	LSATH	-18.0106,15.9949,2.0157,-18.0106	+H	510.267097
5780	LSATH	-18.0106	+H	510.267097
5781	LSATH	-18.0106,2.0157,-2.0157	+H	510.267097
5782	HLSAT	2.0157,-18.0106,-2.0157	+H	510.267097
5783	LSATH	2.0157,-18.0106,-2.0157	+H	510.267097
8986	HLSATH	2.0157,-18.0106,-2.0157	+H	647.326009
8987	HLSATH	-2.0157,2.0157,-18.0106	+H	647.326009
8988	HLSATH	-2.0157,-18.0106,2.0157	+H	647.326009
8989	HLSATH	-18.0106	+H	647.326009
8990	HLSATH	-18.0106,2.0157,-2.0157	+H	647.326009
8991	HLSATH	-18.0106,15.9949,2.0157,-18.0106	+H	647.326009

Figure 88: Output from RiPP peptide mass calculator, after entering masses identified from metabolomic screening alongside putative mass changes and the *S. albus* J1074 core peptide sequence.

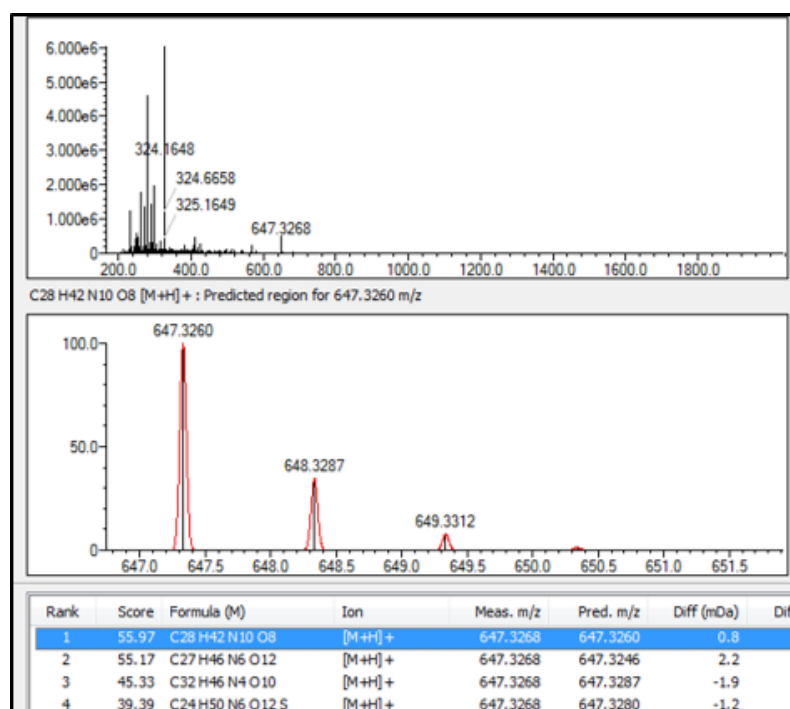


Figure 89: Screenshot of output of Shimadzu formula prediction tool, when analysing the putative final metabolite *m/z* 647.32.

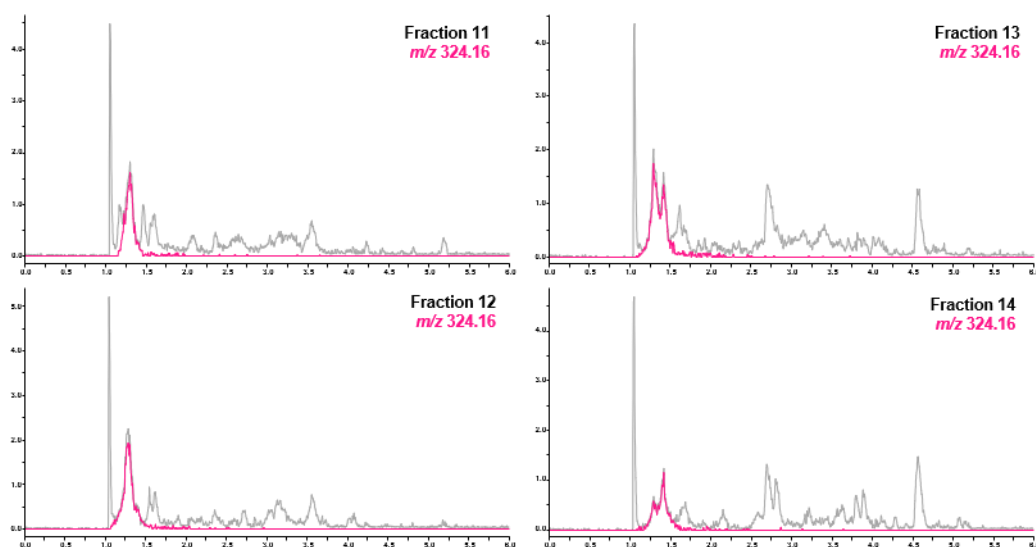


Figure 90: Fractions from Biotage purification that contain the target metabolite.

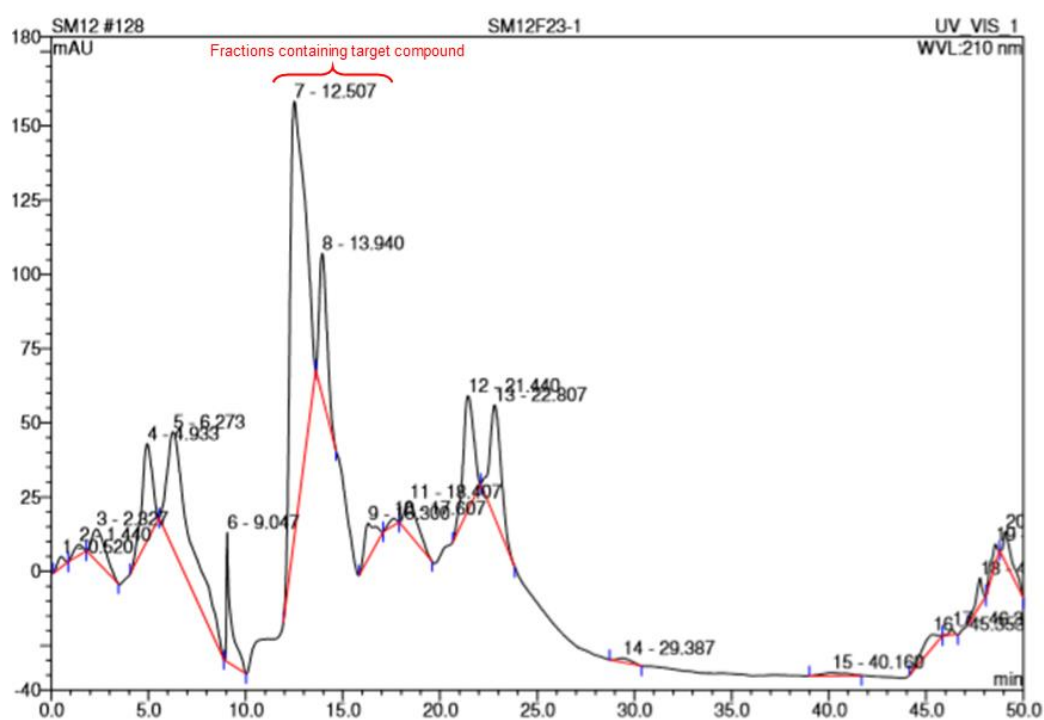


Figure 91: UV chromatogram obtained from semi-prep HPLC purification of fractions containing target metabolite.

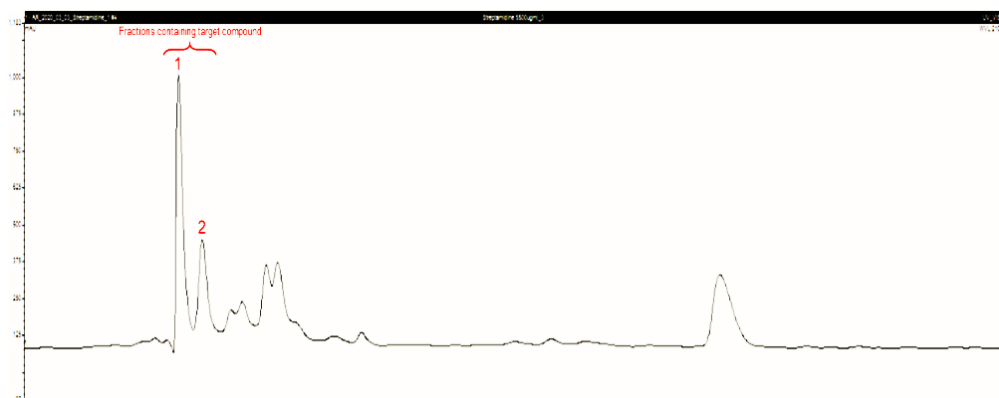


Figure 92: UV chromatogram obtained following semi-prep HPLC purification of semi-pure material from previous HPLC purification step. Samples were subjected to a 10% methanol isocratic gradient. The target metabolite eluted as two distinct peaks indicated on the chromatogram, which were separately collected.

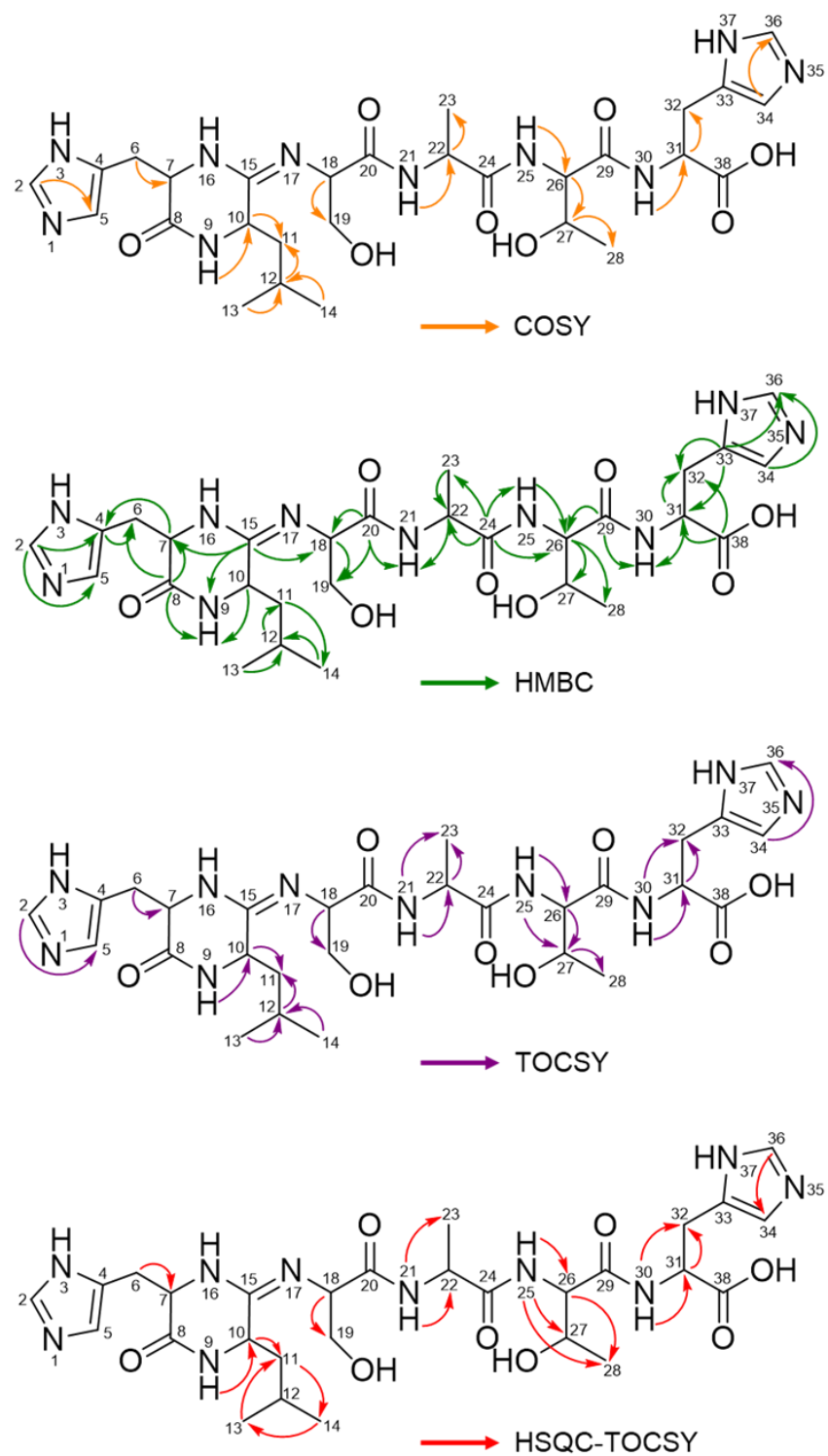


Figure 93: 2D NMR correlations used to elucidate the structure of streptomidine. Coloured arrows represent observed correlations from different NMR experiments: COSY (orange), HMBC (green), TOCSY (purple), HSQC-TOCSY (red).

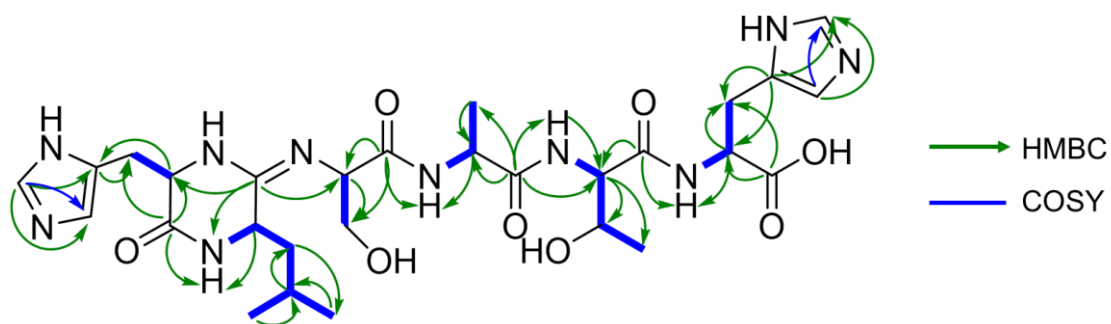


Figure 94: Structure of streptomycin with HMBC and COSY NMR correlations shown in green arrows (HMBC) and blue lines (COSY).

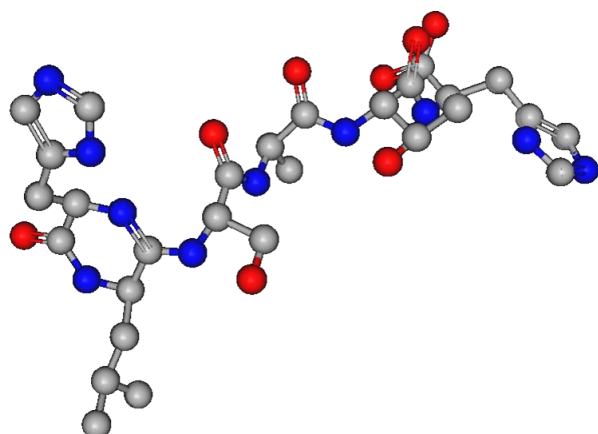


Figure 95: 3D model of streptomycin

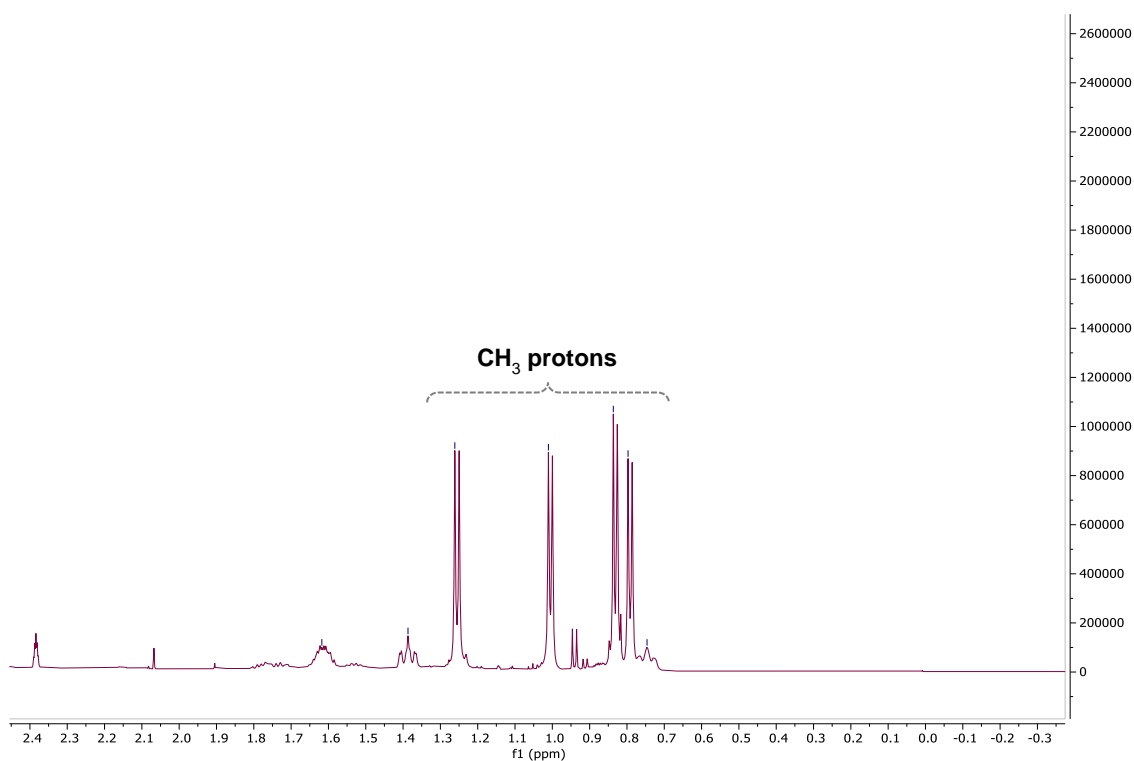


Figure 96: Chemical shifts for CH₃ groups in streptomycin recorded in DMSO-d₆.

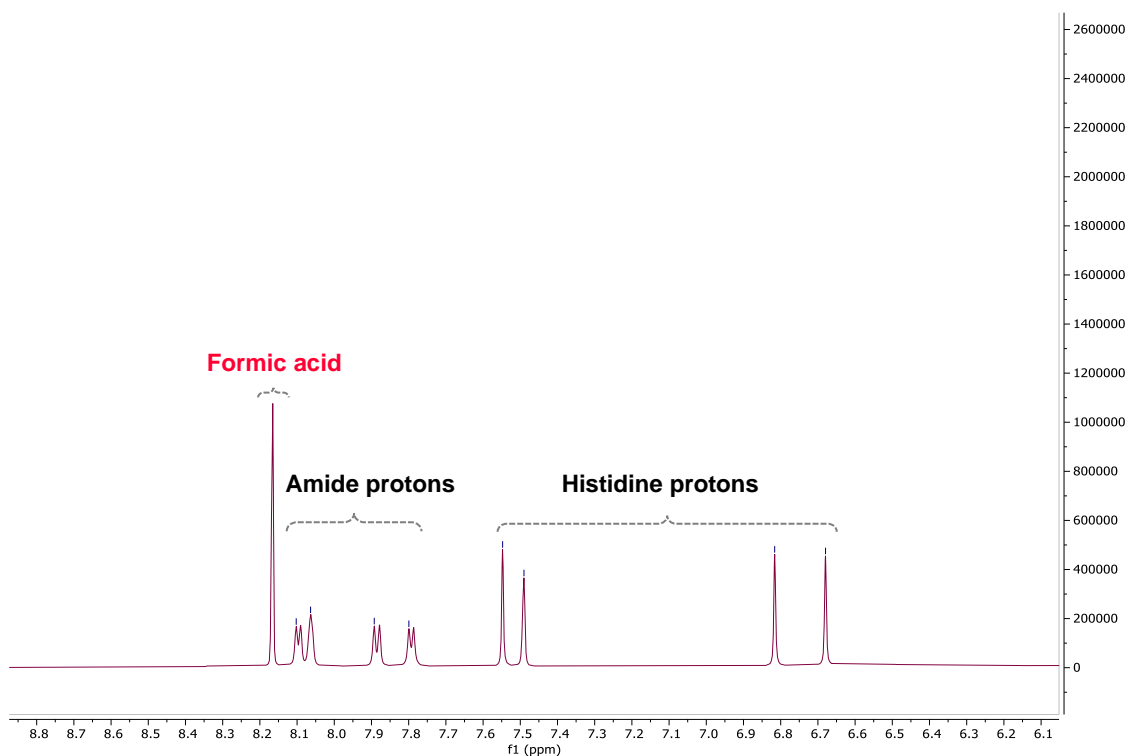


Figure 97: Amide and histidine proton signals for streptamidine recorded in DMSO-d₆.

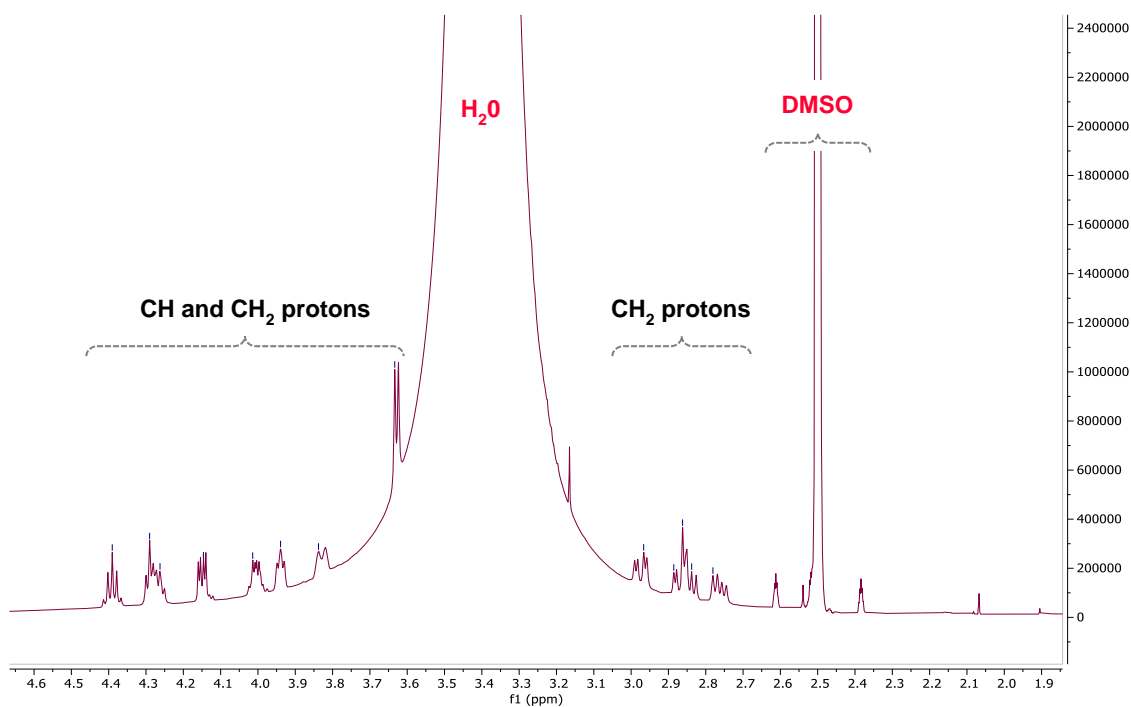


Figure 98: Chemical shifts for CH and CH₂ groups in streptamidine recorded in DMSO-d₆.

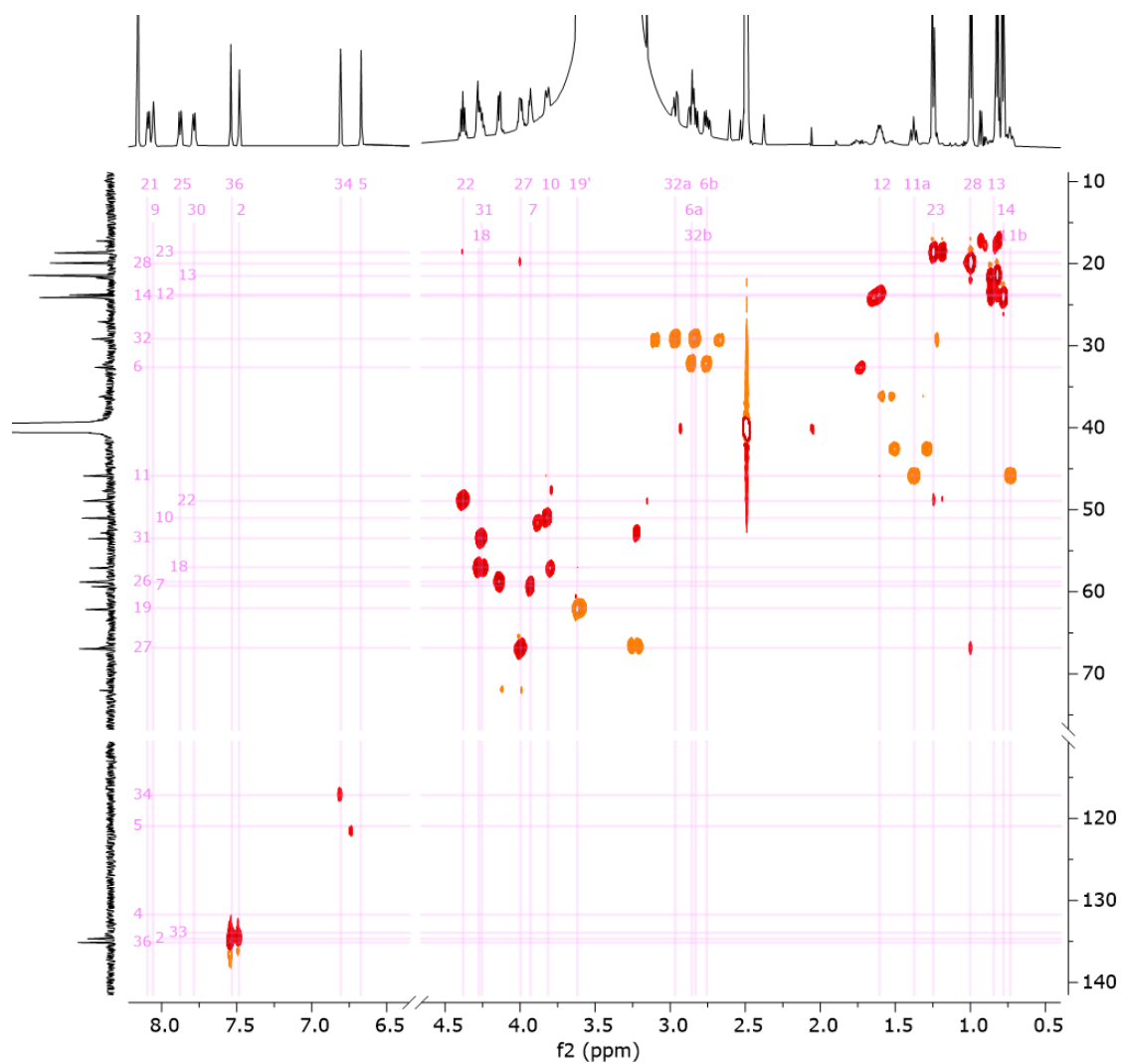


Figure 99: Full HSQC spectrum for streptomidine in DMSO-d₆.

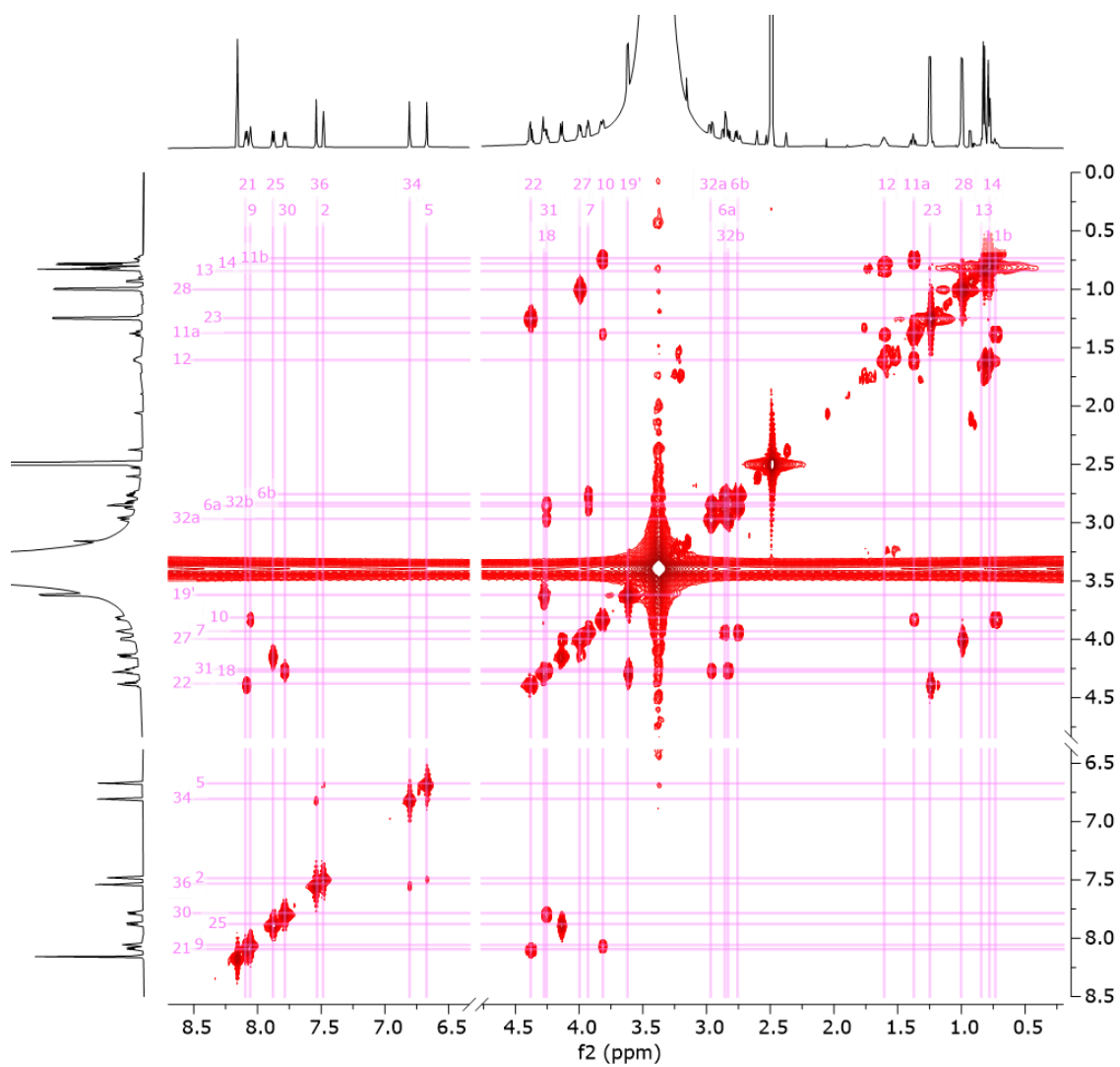


Figure 100: Full COSY spectrum for streptomidine in DMSO-d₆.

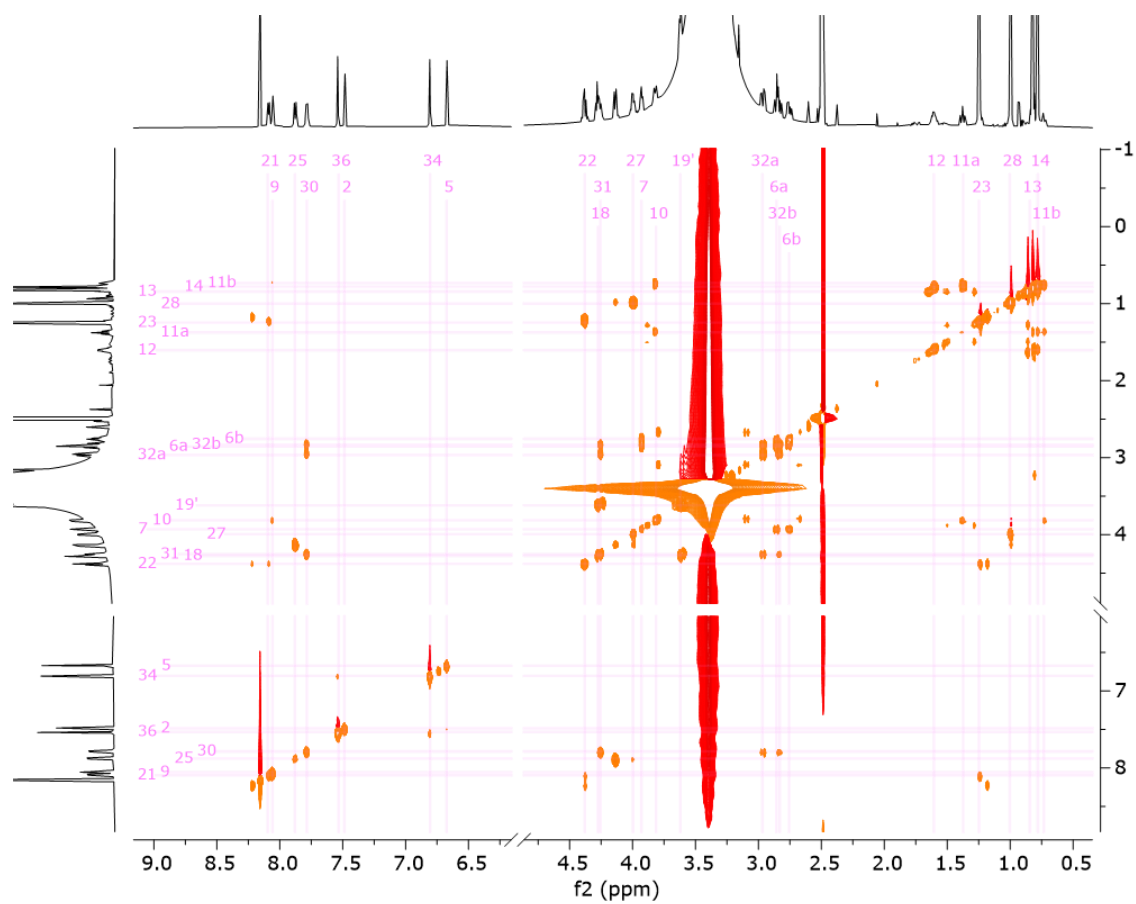


Figure 101: Full TOCSY spectrum for streptomidine recorded in DMSO-d₆.

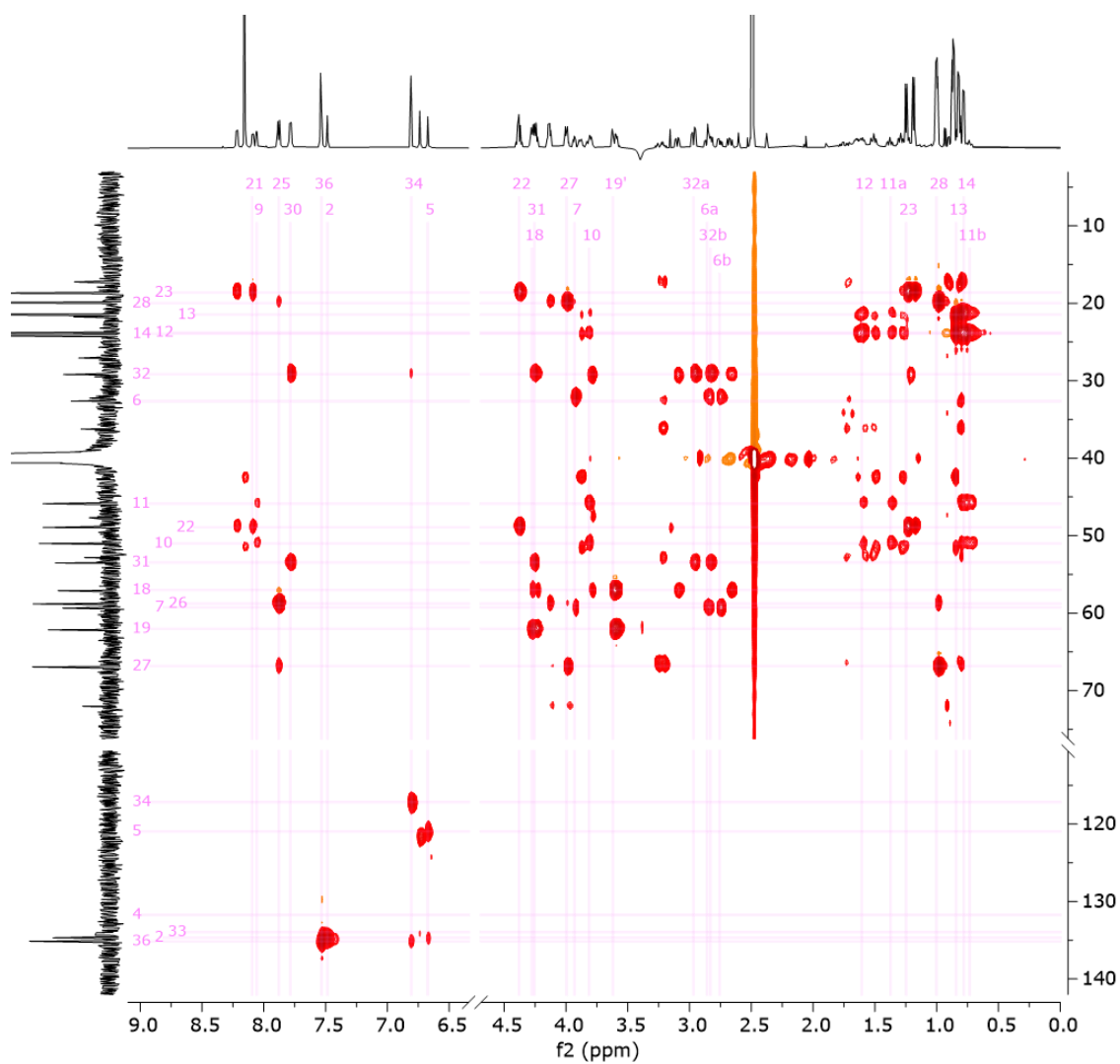


Figure 102: Full HSQC-TOCSY spectrum for DMSO-d₆.

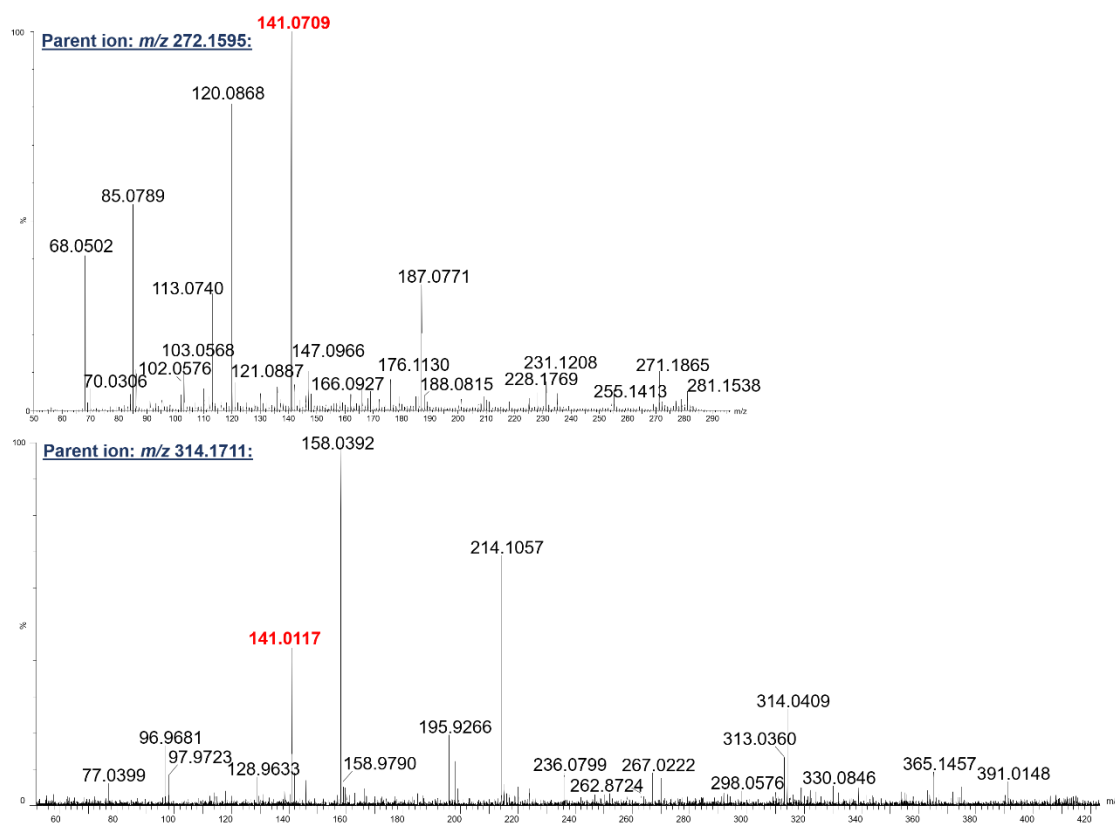


Figure 103: MS² fragmentation data for predicted dehydrated LSA (m/z 272.1595) and predicted acetylated and dehydrated LSA (m/z 314.1711), obtained using a Waters Synapt G2Si.

Table 26: Accurate masses for streptomidine and related metabolites, obtained using a Waters Synapt G2Si.

Description	Formula	Calc. $[M+H]^+$	Obs. m/z	Error (ppm)
Streptomidine (modified HLSATH)	$C_{28}H_{42}N_{10}O_8$	324.1666 ($[M+2H]^{2+}$)	324.1666	0.00
		647.3260	647.3251	1.39
Predicted modified HLSAT	$C_{22}H_{35}N_7O_7$	510.2671	510.2668	0.59
Predicted modified HLSA	$C_{18}H_{28}N_6O_5$	409.2194	409.2195	-0.24
Predicted dehydrated LSA	$C_{12}H_{21}N_3O_4$	272.1604	272.1595	3.31
Predicted acetylated and dehydrated LSA	$C_{14}H_{23}N_3O_5$	314.1710	314.1711	-0.32
Predicted acetylated LSA	$C_{14}H_{25}N_3O_6$	332.1816	332.1818	-0.60

Table 27: Accurate masses for streptomidine and related metabolites, obtained using a Shimadzu IT-TOF.

Description	Formula	Calc. [M+H] ⁺	Obs. <i>m/z</i>	Error (ppm)
Streptomidine (modified HLSATH)	C ₂₈ H ₄₂ N ₁₀ O ₈	647.3260	647.3255	0.77
Predicted modified HLSAT	C ₂₂ H ₃₅ N ₇ O ₇	510.2671	510.2688	-3.33
Predicted modified HLSA	C ₁₈ H ₂₈ N ₆ O ₅	409.2194	409.2175	4.64
Predicted dehydrated LSA	C ₁₂ H ₂₁ N ₃ O ₄	272.1604	314.171	0.00
Predicted acetylated and dehydrated LSA	C ₁₄ H ₂₃ N ₃ O ₅	314.1710	272.1613	-2.94
Predicted acetylated LSA	C ₁₄ H ₂₅ N ₃ O ₆	332.1816	332.1814	0.60

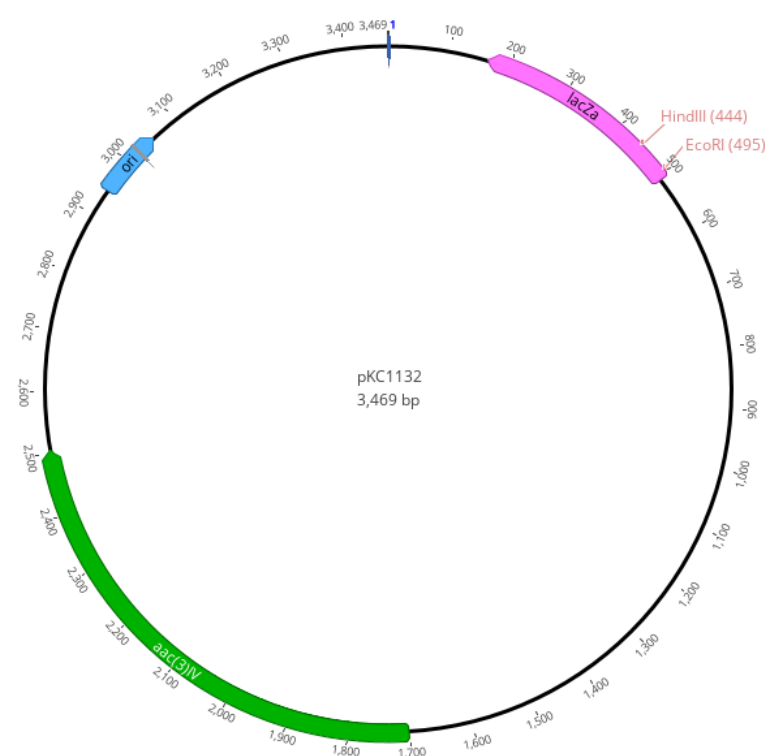


Figure 104: plasmid map of pKC1132, suicide vector used for disruption of the streptomidine biosynthetic gene cluster in *Streptomyces albus* J1074.

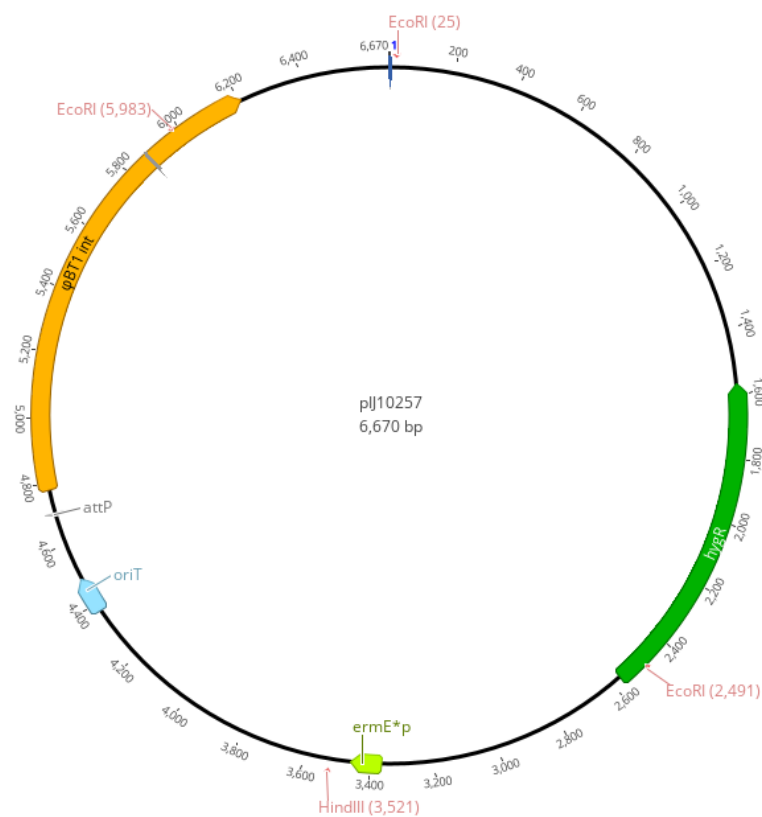


Figure 105: plasmid map for pIJ10257, expression vector used for complementation of gene deletions and overexpression of MarR regulator in *Streptomyces coelicolor* M1146-pCAPSalbC.

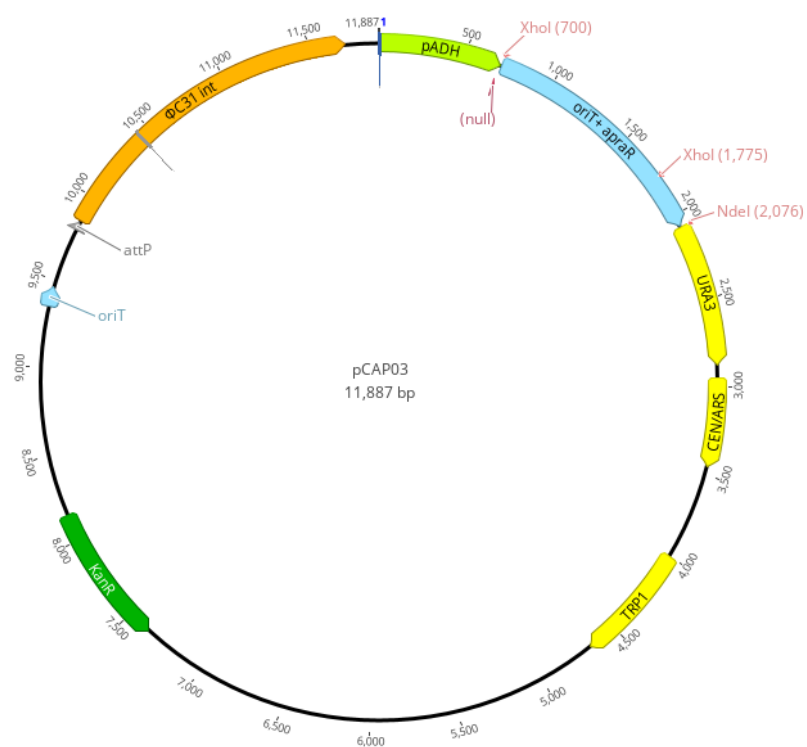


Figure 106: plasmid map for pCAP03, vector used for capture of the streptomidine biosynthetic gene cluster.

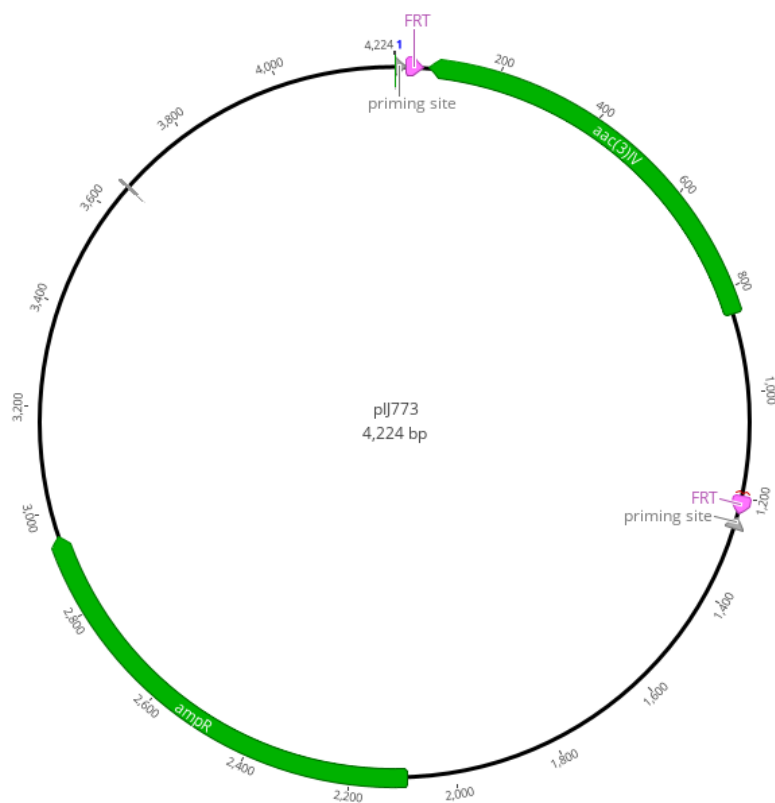


Figure 107: plasmid map for pIJ773 Δ oriT, plasmid used as template for amplification of resistance cassette during PCR targeting experiments.